Inzicht in presterend vermogen van veiligheidsregio's

Onderzoek naar de mogelijkheid en wenselijkheid van een stelsel van indicatoren voor het presterend vermogen van veiligheidsregio's

KWINK GROEP

TU Delft

Inzicht in presterend vermogen van veiligheidsregio's

Onderzoek naar de mogelijkheid en wenselijkheid van een stelsel van indicatoren voor het presterend vermogen van veiligheidsregio's

Delft, 17 / 08 / 2015

Auteurs: Prof.mr.dr. J.A. (Hans) de Bruijn
Dr. M.L.C. (Mark) de Bruijne
Ir. M. (Maarten) Noordink
A. (Anna) Stutje MSc.

Status: Eindrapport
Versie: Definitief

Inhoudsopgave

Samenvatting

Achtergrond en vraagstelling

De commissie Hoekstra heeft in 2013 geconcludeerd dat het toezicht op rampenbestrijding en crisisbeheersing effectiever kan. De Inspectie Veiligheid en Justitie (VenJ), belast met het toezicht op de naleving van de Wet veiligheidsregio's (Wvr), zou zich te veel richten op naleving van de wettelijke normen en te weinig op het feitelijk functioneren van de veiligheidsregio's. Het rapport van de commissie Hoekstra is aanleiding geweest voor de Inspectie VenJ en de Nationaal Coördinator Terrorismebestrijding en Veiligheid (NCTV) om een onderzoek uit te zetten onder leiding van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC). Met het onderzoek is beoogd inzicht te krijgen in de mogelijkheid en wenselijkheid om de prestaties van veiligheidsregio's meer te baseren op output- of zelfs outcome-indicatoren, tegenover de randvoorwaardelijke indicatoren (naleving wettelijke normen, organisatie-eisen) waarop ze nu beoordeeld worden. In het onderzoek staan de volgende hoofdvragen centraal:

- In hoeverre en hoe – op basis van theoretische en praktische inzichten – is het mogelijk om een stelsel van indicatoren te ontwikkelen waarmee een beeld kan worden verkregen van het presterend vermogen van veiligheidsregio's?
- In hoeverre is de ontwikkeling van een stelsel van indicatoren wenselijk vanuit het perspectief van de betrokken actoren?

Aanpak

Het onderzoek is gestart met een literatuurstudie naar de theorie van prestatiemeting en het voeren van acht verkennende gesprekken met zowel afgevaardigden van de Inspectie en de NCTV als directeuren van verschillende veiligheidsregio's. Tevens is documentonderzoek gedaan naar ervaringen met prestatiemeting in het bredere veiligheidsdomein en vijf andere sectoren (zorg, universiteiten, politie, onderhoud civiele infrastructuur en offshore gas). Op basis van de inzichten uit de literatuur en uit de praktijk zijn ontwerpeisen opgesteld. De ontwerpeisen zijn toegepast op de context van de veiligheidsregio's. Hiertoe zijn in de volle breedte van het domein van de veiligheidsregio gesprekken gevoerd, zowel op rijksniveau als op decentraal niveau, en zowel op bestuurlijk niveau als op directieniveau en operationeel niveau. In drie veiligheidsregio's – Amsterdam-Amstelland, IJsselland en Gelderland-Zuid – zijn gesprekken gevoerd en in één regio – Rotterdam Rijnmond – is een systeemoefening bijgewoond. Uit deze gesprekken zijn dominante beelden gedestilleerd, die zijn getoetst in een bijeenkomst met afgevaardigden uit een aantal andere veiligheidsregio's. Dit leidde tot bevestiging van de beelden.

Prestatiemeting in literatuur

Literatuuronderzoek heeft een aantal inzichten opgeleverd die relevant zijn voor de zoektocht naar output- en outcome-indicatoren voor de veiligheidsregio's.

Een goed functionerend systeem van prestatiemeting dient indicatoren te bevatten die hoog scoren op de volgende drie aspecten: **meetbaarheid**, **relevantie** en **toerekenbaarheid**. Meetbaarheid heeft betrekking op de vraag of de prestatie eenduidig te meten is. En dus of aantoonbaar kan worden gemaakt dat sprake is van een verbetering of een verslechtering. Bij relevantie draait het om de vraag of de prestatie uitdrukking geeft aan het bereiken van de doelstelling van de organisatie. Met andere woorden: doet de gemeten prestatie ertoe of niet? Toerekenbaarheid heeft betrekking op de vraag of de prestatie die wordt gemeten toerekenbaar is aan degene die de prestatie op papier levert.

Of prestatiemeting leidt tot positieve effecten of tot perverse effecten hangt ook af van de **impact** van een prestatiemeetsysteem. Is die hoog of laag? Een hoge impact betekent dat het wel of niet behalen van een te meten prestatie grote gevolgen heeft in termen van verantwoording. Het gaat daarbij niet alleen om de feitelijke maar juist ook om de gepercipieerde impact: heeft het wel of niet behalen van een prestatie in de perceptie van de betrokken spelers een hoge impact? Ervaart de professional het systeem als oneerlijk of 'dom' dan neemt de kans op pervertering toe.

Als er sprake is van problematische meetbaarheid, relevantie en toerekenbaarheid in combinatie met een hoge (gepercipieerde) impact, dan is het risico van pervertering zeer groot en zal de sturende werking gering zijn. Het logische alternatief lijkt: reken niet hard af op prestaties, maar richt prestatiemeting op 'leren'. Het risico is dan echter dat er weinig sturende werking van prestatiemeting uitgaat. De effectiviteit van prestatiemeting is optimaal als een goede balans wordt gevonden in de sterkte van de prikkels doordat zowel aandacht is voor 'leren' als 'verantwoorden'.

Prestatiemeting in praktijk

We hebben de ervaringen in de sectoren zorg, universiteiten, politie, onderhoud civiele infrastructuur en off-shore gas onderzocht en ook eerdere initiatieven met prestatie-indicatoren in het veiligheidsdomein bestudeerd. Dat leverde inzichten op voor de zoektocht naar output- en outcome-indicatoren voor de veiligheidsregio's:

- Prestatiemeting en de bijbehorende stelsels zijn complex. Er is sprake van een veelheid aan actoren en functies die aan prestatiemeting worden gekoppeld. De toerekenbaarheid van prestaties aan actoren vormt veelal een probleem.
- In geen van de onderzochte sectoren zien we het gebruik van betekenisvolle outcome-indicatoren, er is sprake van hooguit output-indicatoren.
- Ervaringen met eerdere initiatieven op het gebied van prestatiemeting in het veiligheidsdomein (bijvoorbeeld Aristoteles en RemBrand) laten zien dat een zuiver kwantitatieve benadering van prestatie-indicatoren niet tot de gewenste resultaten leidt en 'schijnzekerheid' creëert.
- Als de indicatoren in de prestatiemeetsystemen een hoge impact hebben, is het risico op perverse effecten eveneens hoog.
- In veel van de onderzochte sectoren zijn systemen 'doorontwikkeld'. Het lijkt er bijvoorbeeld op dat de aandacht verschuift van scores op indicatoren naar een gesprek met behulp van indicatoren.

Ontwerpeisen

De inzichten uit de praktijk en theorie hebben wij vertaald in een aantal ontwerpeisen. De gedachte achter de ontwerpeisen is dat een stelsel van indicatoren ten minste aan deze eisen zal moeten voldoen, wil het betekenisvol en kansrijk zijn, en wil voorkomen worden dat de toepassing ervan leidt tot perverse effecten (bijvoorbeeld 'afvinken'). In het onderzoek zijn de volgende vier ontwerpeisen geformuleerd:

1. De indicatoren dienen zoveel mogelijk **meetbaar** te zijn.
2. De indicatoren dienen zo **relevant** mogelijk te zijn.
3. De indicatoren dienen zoveel mogelijk **toerekenbaar** te zijn.
4. Naarmate meetbaarheid, relevantie en toerekenbaarheid meer problematisch zijn, dient de **gepercipieerde impact** van het prestatiemeetsysteem in de veiligheidsregio's beperkter te zijn om perverse effecten te voorkomen.

Beelden uit de regio's over prestatiemeting

Uit de gesprekken in de regio's hebben we dominante beelden gedestilleerd. De beelden hebben betrekking op de wijze waarop in de veiligheidsregio's tegen nut en noodzaak van prestatiemeting wordt aangekeken.

De onderzochte regio's erkennen het belang van inzicht in prestaties, van de systeemverantwoordelijkheid van de minister en van verticale verantwoording. Gesprekspartners in de regio's erkennen ook het belang van prestatie-indicatoren, daar waar het gaat om de wettelijke vereisten. Tegelijkertijd is er kritiek op, dan wel angst voor een mechanistisch gebruik van prestatie-indicatoren. Er is weinig steun voor het vervangen van huidige indicatoren door output- en outcome-indicatoren. Ten eerste omdat output- en outcome-prestaties in veiligheidsregio's zich moeilijk laten vatten in een systeem van indicatoren. Ten tweede omdat door de lage frequentie en het unieke karakter van een ramp/crisis de zeggingskracht van indicatoren in de warme fase beperkt is. Ten derde omdat de relatie koud – warm in een systeem van prestatiemeting problematisch is. Ten vierde omdat er veel angst is dat indicatoren zullen worden gebruikt voor vormen van afrekenen – in de brede zin van het woord (bijvoorbeeld door te ranken). Indicatoren hebben een gepercipieerde hoge impact.

Er zijn ons ook andere zienswijzen aangereikt, deels vanuit de regio en deels vanuit de Inspectie en de NCTV. Zo kwam de roep om betere indicatoren dan de huidige proces-indicatoren ook vanuit de regio zelf (Veiligheidsberaad). Hierdoor ontstaat een diffuus beeld: enerzijds is er vanuit de regio de roep om betere indicatoren, anderzijds wordt dit desgevraagd niet concreet gemaakt en wordt door regio's vooral gewezen op de nadelen van indicatoren. Ook bij de kritiek vanuit de regio op het ranken van veiligheidsregio's is een andere zienswijze mogelijk. Rankings hebben op z'n minst bestuurders geactiveerd en zo tot activiteit en beweging geleid. De drang om in het 'linkerrijtje' terecht te komen, kan ook positief worden geduid.

Conclusies

Na toepassing van de ontwerpeisen concluderen wij dat het – bezien vanuit de theorie en de praktijk – niet mogelijk en wenselijk is om met een stelsel van output- en/of outcome-indicatoren uitdrukking te geven aan het presterend vermogen van veiligheidsregio's.

- Ten eerste zijn meetbaarheid en toerekenbaarheid problematisch. In het onderzoek genoemde voorbeelden van output- en outcome-indicatoren zijn in veel gevallen weliswaar relevant, maar zelden goed meetbaar en toerekenbaar.
- Ten tweede kennen veiligheidsregio's een aantal bijzonderheden die het problematische karakter van het gebruik van output- en outcome-indicatoren versterken:
    - *De koud-warm dynamiek.* Prestaties in de 'warme'-fase (als zich daadwerkelijk een incident voordoet) hebben de meeste zeggingskracht, maar zijn uniek en komen weinig voor en zijn daardoor niet te vangen in indicatoren. Prestaties in de 'koude'-fase (als er niets aan de hand is en de nadruk ligt op zaken als opleiden en oefenen) zijn veelvoorkomend, maar niet in output- en outcome-indicatoren te vangen.
    - *Het samenspel van mono en multi.* Er zijn kolommen (mono) en er is de regio die toegevoegde waarde heeft ten opzichte van de kolommen (multi). Veel van de activiteiten van de regio zijn relationeel van aard: overleg, samenwerking, kennisuitwisseling, afstemming en evalueren. Die kun je hooguit in procesindicatoren vangen, die weinig zeggingskracht hebben.
- Ten derde zijn wij in de regio's gestuit op veel weerstand tegen (extra) indicatoren. Hierbij zijn allerlei overwegingen genoemd: de variëteit tussen de regio's (die op gespannen voet staat met een eenduidig systeem), het relatief korte bestaan van de regio's (die nog in ontwikkeling zijn) en de spanning tussen verticale en horizontale verantwoording (die ongelukkig interfereren). Ook al zou er geen weerstand zijn, feit is en blijft dat meetbaarheid en toerekenbaarheid problematisch zijn – en dus de ontwikkeling van een set van output- en outcome-indicatoren in de weg staan. Daarnaast is er het risico dat bij zo

veel wantrouwen tegen dit soort indicatoren, de perceptie ontstaat dat deze een te hoge impact gaan krijgen – hetgeen een prikkel voor pervertering kan zijn.

De combinatie van problematische meetbaarheid en toerekenbaarheid en hoge gepercipieerde impact leidt tot sterke prikkels voor pervertering en ritualisering. De sturende werking van een systeem bestaande uit output- en outcome-indicatoren zal gering zijn. Het zal – net als het huidige systeem – leiden tot 'afvinken' en een schijnwerkelijkheid creëren.

Reflectie

De conclusie dat het ontwikkelen en implementeren van een stelsel van output- en outcome-indicatoren een weinig heilzame weg is, roept de vraag op hoe de Inspectie VenJ en de NCTV dan wel beter inzicht kunnen verkrijgen in het presterend vermogen van veiligheidsregio's. Deze behoefte is immers volstrekt legitiem.

Wij beargumenteren dat wanneer wordt erkend dat het werk van de regio's zich niet in een stelsel van indicatoren laat vangen, dit wel eens de ruimte zou kunnen bieden voor veel vruchtbaarder discussies over verantwoording en leren. De aandacht zou moeten verschuiven van een focus op output- en outcome-indicatoren naar het 'gesprek' of de 'kritische interactie' tussen het Ministerie van VenJ (Inspectie VenJ en NCTV) en regio. Dat biedt de mogelijkheid om tot de essentie van waar het om gaat in de veiligheidsregio's door te dringen.

De aandacht zou daarmee moeten verschuiven van een zoektocht naar output- en outcome-indicatoren naar de vormgeving van die kritische interactie – naar de vraag wat (1) een goed inhoudelijk format is voor zo'n gesprek, (2) de spelregels zijn voor zo'n gesprek en (3) hoe wordt omgegaan met de spanning tussen verantwoorden en leren.

Van indicatoren naar een gesprek: het lijkt de overgang van iets dat hard en uitdagend is (indicatoren) naar iets dat zacht en (te) gemakkelijk is (gesprek). Het tegenovergestelde is waar. Indicatoren zijn bij nadere beschouwing vaak boterzacht. En een gesprek kan een harde confrontatie zijn. Bovendien, een gesprek vergt ook wat van de Inspectie en de regio. Indicatoren scheppen afstand: je meet of een indicator is behaald en komt tot je conclusie. Daarvoor is, bij wijze van spreken, geen enkel onderling contact nodig. Als de aandacht verschuift naar het gesprek, is interactie geboden. De regio moet zich intelligenter verantwoorden. De Inspectie moet intelligenter kijken.

Summary

Background and remit

In 2013, the Hoekstra Committee concluded that supervision of disaster and crisis prevention could be made more effective. The Committee indicated that the Security and Justice Inspectorate, which is charged with monitoring compliance with the Safety Regions Act (Wvr), focuses too much on compliance with legal requirements and too little on the actual functioning of safety regions. The report of the Hoekstra Committee prompted the Security and Justice Inspectorate and the National Coordinator for Security and Counterterrorism (NCTV) to arrange for further research to be conducted under the guidance of the Scientific Research and Documentation Centre (WODC). The aim of this research was to gain insight into the feasibility and desirability of assessing the performance of safety regions on the basis of more output indicators or even outcome indicators, as compared to the current indicators based on parameters (compliance with legal standards, organisational requirements). The study focused on the following questions:

- To what extent and how, on the basis of theoretical and practical knowledge, might it be possible to develop a set of indicators that could capture the performance capacity of the safety regions?
- To what extent would the development of a system of indicators be desirable from the stakeholders' perspective?

Approach

The study began with a literature study on the theory relating to performance. In addition, eight exploratory meetings were held with representatives from both the Inspectorate and the NCTV, which are responsible for the management of the various safety regions. Documentary research was also carried out to chart experiences with performance measurement in the wider field of security and five other sectors (healthcare, universities, police, civil infrastructure maintenance and offshore gas). Design criteria were drawn up, based on insights from both the literature and practice, and applied in the context of the safety regions. For this purpose, meetings were held across the full breadth of the safety region domain, at the national and local levels, and at the administrative, management and operational levels. Discussions were held in three safety regions (Amsterdam-Amstelland, IJsselland and Gelderland-Zuid) and a system exercise was attended in one region (Rotterdam Rijnmond). These discussions produced a range of views on how the safety regions should be assessed and the need for and utility of performance measures. The impressions that emerged from these discussions were tested at a meeting with representatives of a number of other safety regions, confirming the views that had been expressed.

Performance measurement in the literature

A study of the literature provided several insights of relevance to the search for output and outcome indicators for the safety regions.

An effective performance measurement system must include indicators that score highly in relation to the following three aspects. **Measurability, relevance** and **accountability**. The concept of 'measurability' relates to the question of whether performance can be measured unambiguously, and therefore whether it can be demonstrated that an improvement or deterioration has taken place. Relevance centres on the question of whether the performance will help to achieve the organisation's goal. In other words, does the performance measured make any difference to the organisation? Accountability relates to whether the performance

measured can be attributed to a particular individual, who is formally accountable for performance in that area.

Whether the measurement of performance has beneficial or adverse effects is partly dependent on the **impact** of a performance measurement system. Is the impact high or low? A higher impact means that whether a performance measure is attained or not will have major implications in terms of holding people to account. At stake here is not only the actual impact, but also the perceived impact: in the view of the players involved, has attaining a good level of performance had a significant impact? If professionals see the system as unfair or 'stupid', there is a higher likelihood of misuse of the indicators.

If problems with measurability, relevance and accountability coincide with a significant (perceived) impact, the risk of misuse is extremely high and the incentive effect will be limited. The logical alternative would seem to be to focus performance measurement on 'learning' rather than focusing strictly on performance. However, this runs the risk of performance measurement having only a limited incentive effect. Performance measurement is most effective when a good balance is struck in terms of the strength of incentives by focusing both on 'learning' and 'accountability'.

Performance measurement in practice

We investigated experiences in the healthcare, university, police, maintenance, civil infrastructure and offshore gas sectors, as well as studying previous initiatives involving performance indicators in the safety domain. This provided insights in relation to the search for output and outcome indicators for the safety regions:

- Performance measurement and the systems associated with it are complex. There are multiple actors and functions linked to performance measurement. Holding actors to account for their performance can often be problematic.
- We did not encounter the use of significant outcome indicators in any of the sectors investigated; at best, what we saw were output indicators.
- Experience of previous initiatives involving performance measurement in the safety domain (e.g., Aristoteles and RemBrand) shows that a purely quantitative approach to performance indicators fails to achieve the desired results and creates a false sense of security.
- If the indicators in performance management systems have a large impact, the risk of perverse effects will also be significant.
- In many of the sectors studied, systems have been 'further developed'. For example, it appears that the focus has shifted from scores based on indicators to a discussion aided by the use of indicators.

Design requirements

The insights from practice and literature are translated in design requirements. The design requirements are based on the idea that at a minimum, the set of indicators should meet these requirements in order to be meaningful and successful, and in order to prevent indicators being used in ways that produce unintended adverse effects (e.g. 'box-ticking'). The following four design requirements were formulated:

1. The indicators in the system should be **measurable** wherever possible.
2. The indicators should be as **relevant** as possible.
3. The indicators should be **accountable** wherever possible.
4. Although measurability, relevance and accountability are important issues, the **perceived impact** of the performance measurement system for the safety regions should be limited in order to avoid unintended adverse effects.

Impressions of performance measurement from the regions

We have gathered the dominant impressions from discussions in the regions. These impressions concern how the necessity and usefulness of performance measures are seen in the safety regions.

The regions studied acknowledge the importance of information on performance, the minister's responsibility for the system and vertical accountability. Those interviewed in the regions also acknowledged the importance of performance indicators when these concern statutory requirements. However, we also encountered criticism and fear of the mechanical use of performance indicators. There is little enthusiasm for the replacement of the existing indicators by output and outcome indicators. First, this is because output and outcome performance in safety regions is difficult to express in a system of indicators. Second, it is because the low frequency and unique character of a disaster/crisis mean that the validity of indicators is limited in the hot phase (when an incident actually occurs). Third, it is because the relationship between hot and cold in a performance measurement system is problematic. Fourth, it is because there is great fear that indicators will be used for other forms of accountability, in the broadest sense of the word (for example, by means of ranking). Indicators are perceived to have a high impact.

Alternative views were also expressed, partly from the regions and partly from the Inspectorate and the NCTV. For example, the call for indicators that are better than the current process indicators also came from the region itself (Safety Advisory Board/*Veiligheidsberaad*). This creates a confused picture: on the one hand, there are calls for better indicators from the region, and on the other, no further specification was provided when requested and the regions primarily highlight the disadvantages of indicators. The region's criticism of the ranking of safety regions can also be seen from a different perspective. At the very least, rankings have prompted policymakers to take action. The pressure to move higher in the rankings can also be interpreted in a positive light.

Conclusions

After applying the design requirements and having considered theory and practice, we conclude that it is neither possible nor desirable to apply a system of output and/or outcome indicators to assess the performance capacity of the safety regions.

- First, measurability and accountability are problematic issues. Although the examples of output and outcome indicators mentioned in the study are certainly relevant in many cases, they are seldom measurable and accountable.
- Second, safety regions have a number of characteristics that reinforce the problematic nature of the use of output and outcome indicators:
  - The hot-cold dynamic. Performance in the 'hot' phase is the most significant, but incidents are unique and rare occurrences; it is therefore not possible to assess performance in this phase using indicators. Performance in the 'cold' phase (when no incident has occurred and the focus is on issues such as training and exercises) can be measured much more easily, but it is impossible to assess performance in this phase using output and outcome indicators.
  - Mono/multi-interplay: the interaction between the separate vertical columns (mono) and the region that is meant to provide added value by linking these columns (multi). Many of the activities in the regions are inter-relational in nature (consultation, cooperation, knowledge-sharing, coordination and evaluation). The only way to assess these is by using process indicators, but these provide little information.
- Third, we encountered great resistance in the regions to the very idea of indicators or additional indicators. A range of reasons was given for this: regional diversity (which is at odds with a unified system), the relatively short history of the regions (which are still developing) and the tension between vertical and horizontal accountability (which creates unintended interference). Even without this

resistance, however, it would remain the case that measurability and accountability are problem areas that would obstruct the development of a set of output and outcome indicators. There is also the risk that mistrust of these indicators could give rise to the perception that they would have an excessive impact, which could create an incentive to misuse the indicators.

The combination of problematic measurability and accountability and the perception that the indicators would have a high impact could create strong incentives for misusing the indicators. A system consisting of output and outcome indicators would have a limited incentive effect. Indeed, it would to lead to a 'box-ticking' approach and create the illusion of performance, as is the case with the current system.

Reflection

The conclusion that it would be inadvisable to develop and implement of a system of output and outcome indicators raises the question as to how the Inspectorate and the NCTV might improve oversight of the performance capacity of the safety regions. After all, it is perfectly legitimate for these institutions to seek such oversight.

We would argue that once it has been acknowledged that it is impossible to evaluate the work of the regions using a system of indicators, this might create the space needed for further productive discussions about accountability and learning. A shift is required, away from identifying new output and outcome indicators and towards 'dialogue' or 'critical interaction' between the government (the Inspectorate and the NCTV) and the regions. This would also create an opportunity to identify what really matters in the safety regions.

In order for this shift to occur, attention would need to move away from the search for output and outcome indicators and towards the design of critical interaction; that is, towards the question of (1) the appropriate substantive format for such a dialogue, (2) the rules that would apply to such a dialogue, and (3) how to deal with the tension between accountability and learning.

Some might see a shift away from indicators and towards dialogue as a transition from the hard and the challenging (indicators) towards something that is soft and easy, perhaps excessively so (dialogue). However, the opposite is true. On closer inspection, indicators are often much softer than they first appear, whereas a dialogue can assume the form of a hard challenge or a confrontation. Dialogue also involves contributions from both the Inspectorate and the region. Indicators create distance: one measures whether an indicator has been achieved and one reaches a conclusion; it could be said that no interaction at all is required. Dialogue, by contrast, involves interaction by definition. It would demand more intelligent answers from the region, and the Inspectorate would have to adopt a more intelligent approach to the issue.

1. Inleiding

1.1. Context

Sinds 1 oktober 2010 is de Wet veiligheidsregio's (Wvr) van kracht. Het doel van de wet is het realiseren van een efficiënte en kwalitatief hoogwaardige organisatie van de brandweerzorg, geneeskundige hulpverlening, rampenbestrijding en crisisbeheersing onder één regionale bestuurlijke regie.[1]

De verantwoordelijkheid voor veiligheid is op gemeentelijk niveau belegd. Het college van B&W is primair verantwoordelijk voor de organisatie van de brandweerzorg, geneeskundige hulpverlening, rampenbestrijding en crisisbeheersing. De burgemeester voert het bevel als zich in zijn of haar gemeente een incident, ramp of crisis voordoet. De voorzitter van de veiligheidsregio heeft de leiding bij rampen of crises die de gemeentegrens overschrijden of meer dan een plaatselijke betekenis hebben.

Binnen de veiligheidsregio werken verschillen organisaties in de regio samen. Voor de aanpak van incidenten hebben zij een gezamenlijke crisisorganisatie ingericht. De organisatie van deze regionale rampenbestrijding en crisisbeheersing is belegd bij de veiligheidsregio, een openbaar lichaam, dat bestuurd wordt door een bestuur van de burgemeesters van de verschillende gemeenten in zijn grondgebied. Dat bestuur heeft een vaste voorzitter, bij koninklijk besluit benoemd uit de burgemeesters van de gemeenten in de regio. De veiligheidsregio wordt bij gemeenschappelijke regeling ingesteld. Het instellen van een veiligheidsregio is verplicht.

De minister van Veiligheid en Justitie (VenJ) is verantwoordelijk voor het stelsel van rampenbestrijding en crisisbeheersing. De minister kan op grond van de Wvr landelijke doelstellingen opleggen. In de wet is vastgelegd dat de voorzitter van de veiligheidsregio de minister een rapportage toestuurt over de uitvoering van de landelijke doelstellingen door de veiligheidsregio. De minister stuurt verder op de verbetering van de brandweerzorg, rampenbestrijding en crisisbeheersing, via onder andere de instelling van het Instituut Fysieke Veiligheid, versterking van het brandweeronderwijs, bovenregionale samenwerking, expertregio's, vitale sectoren en de samenwerking met Defensie.[2] Daarnaast worden er sinds 2012 gezamenlijke prioriteiten met het Veiligheidsberaad vastgesteld.[3] Het Veiligheidsberaad bestaat uit de 25 voorzitters van de veiligheidsregio's.

De Inspectie VenJ is op grond van de Wvr, onder gezag van de minister van VenJ, belast met het toezicht op de taakuitvoering door de veiligheidsregio's. De Inspectie beoordeelt de veiligheidsregio's met het wettelijk kader als uitgangspunt. De Inspectie rapporteert daarover richting de minister o.a. in de Staat van de rampenbestrijding.

[1] Kamerstukken II, 2006/07, 31117, nr. 3, p. 1-2.
[2] Veldhuisen et al., 2013, p. 14.
[3] Veldhuisen et al., 2013, p. 14.

1.2. Achtergrond onderzoek

Op verzoek van de minister van VenJ heeft de onafhankelijke commissie Hoekstra in 2012 en 2013 onderzoek gedaan naar de werking van de Wvr en het brede Nederlandse stelsel van rampenbestrijding en crisisbeheersing. In september 2013 heeft de commissie haar advies aan de regering uitgebracht.[4]

De Commissie Hoekstra heeft geconcludeerd dat het toezicht op rampenbestrijding en crisisbeheersing effectiever kan. De Inspectie VenJ, belast met het toezicht op de naleving van de Wvr zou zich te veel richten op naleving van de wettelijke normen en te weinig op het feitelijk functioneren van de veiligheidsregio's.[5]

Het rapport van de commissie Hoekstra is aanleiding geweest voor de Inspectie en de Nationaal Coördinator Terrorismebestrijding en Veiligheid (NCTV) om een onderzoek uit te zetten onder leiding van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC). Centraal in het onderzoek staat de vraag in hoeverre en op welke wijze het mogelijk is om een stelsel van indicatoren te ontwikkelen waarmee een beeld kan worden verkregen van het presterend vermogen van veiligheidsregio’s, evenals de vraag in hoeverre dit wenselijk is vanuit het perspectief van de betrokken actoren.

1.3. Vraagstelling

Met het onderzoek wordt beoogd inzicht te krijgen in de mogelijkheid en wenselijkheid om de prestaties van veiligheidsregio's meer te baseren op output- of zelfs outcome-indicatoren, tegenover de randvoorwaardelijke indicatoren (naleving wettelijke normen, organisatie-eisen) waarop ze nu beoordeeld worden.

In het onderzoek staan daartoe de volgende hoofdvragen centraal:

- In hoeverre en hoe – op basis van theoretische en praktische inzichten – is het mogelijk om een stelsel van indicatoren te ontwikkelen waarmee een beeld kan worden verkregen van het presterend vermogen van veiligheidsregio’s?
- In hoeverre is de ontwikkeling van een stelsel van indicatoren wenselijk vanuit het perspectief van de betrokken actoren?

Om antwoord te geven op de centrale hoofdvragen zijn onderstaande deelvragen geformuleerd. In de tabel wordt weergeven in welk hoofdstuk de deelvragen aan bod komen.

[4] Evaluatiecommissie Wet veiligheidsregio's en het stelsel van rampenbestrijding en crisisbeheersing (18 september 2013). De commissie heeft haar advies mede gebaseerd op:
- de "Staat van de rampenbestrijding" van de Inspectie VenJ met daarin de operationele prestaties van de veiligheidsregio's;
- de in opdracht van het WODC uitgevoerde evaluatie van de Wvr over het functioneren van het stelsel krachtens de Wvr en hoe actoren dat ervaren.

[5] Evaluatiecommissie Wet veiligheidsregio's en het stelsel van rampenbestrijding en crisisbeheersing (18 september 2013).

Onderzoeksvragen	Hoofdstuk
1. In hoeverre zijn, bezien vanuit een beknopte vogelvlucht, prestatie-indicatoren in theorie een geschikt instrument om een beeld te vormen van het presterend vermogen van veiligheidsregio's?	Hoofdstuk 2 en 6
2. In hoeverre is het gebruik van prestatie-indicatoren voor het presterend vermogen van veiligheidsregio's vanuit praktisch oogpunt mogelijk? Om welke indicatoren gaat het, hoe kan op basis van die indicatoren een oordeel worden gegeven over het presterend vermogen van veiligheidsregio's, hoe kunnen negatieve/perverse effecten van prestatiemeting vermeden worden?	Hoofdstuk 4 en 6
3. Welke ervaringen hebben andere actoren met (output- en vooral op outcome gerichte) prestatie-indicatoren: • Welke ervaringen met prestatie-indicatoren bestaan er binnen de veiligheidswereld (bijvoorbeeld RemBrand en Aristoteles/Cicero)? • Welke ervaringen met prestatie-indicatoren hebben andere toezichthouders dan de Inspectie VenJ? • Welke ervaringen met prestatie-indicatoren hebben actoren vanuit andere domeinen binnen de overheid? • Welke ervaringen met prestatie-indicatoren zijn er opgedaan in het bedrijfsleven (bijvoorbeeld de dienstensector)?	Hoofdstuk 3
4. Hoe staan bestuurders en operationeel verantwoordelijken binnen de veiligheidswereld ten opzichte van prestatie-indicatoren: • Welke (elementen/aspecten van) prestatie-indicatoren wil men hanteren om te kunnen sturen op het eigen functioneren en waarom die indicatoren; welke niet en waarom niet? • Wat zijn volgens bestuurders en operationeel verantwoordelijken de prestatie-indicatoren waarop de veiligheidsregio's beoordeeld moeten worden en waarom die indicatoren; welke niet en waarom niet?	Hoofdstuk 5
5. Welke kritische succesfactoren zijn er op basis van eerdere incidentonderzoeken te formuleren voor het bepalen van het presterend vermogen van veiligheidsregio's?	Hoofdstuk 3
6. In hoeverre en hoe kunnen de theoretische en praktische inzichten vanuit de voorgaande onderzoeksvragen worden vertaald in een stelsel van indicatoren (of zelfs een meetlat/instrument) waarmee een beeld kan worden verkregen van het presterend vermogen van veiligheidsregio's? In hoeverre kan dat niet en waarom niet?	Hoofdstuk 7

Accenten in het onderzoek

In overleg met de begeleidingscommissie is ervoor gekozen om het onderzoek af te bakenen tot de taak van de veiligheidsregio's om kolom overstijgende rampenbestrijding en crisisbeheersing te organiseren. Hoewel de veiligheidsregio ook verantwoordelijk is voor de dagdagelijkse taken op het terrein van brandweerzorg en de geneeskundige hulpverleningsorganisatie in de regio (GHOR), staat in dit onderzoek het multidisciplinaire karakter van de veiligheidsregio's centraal in opgeschaalde situaties en de voorbereiding daarop.

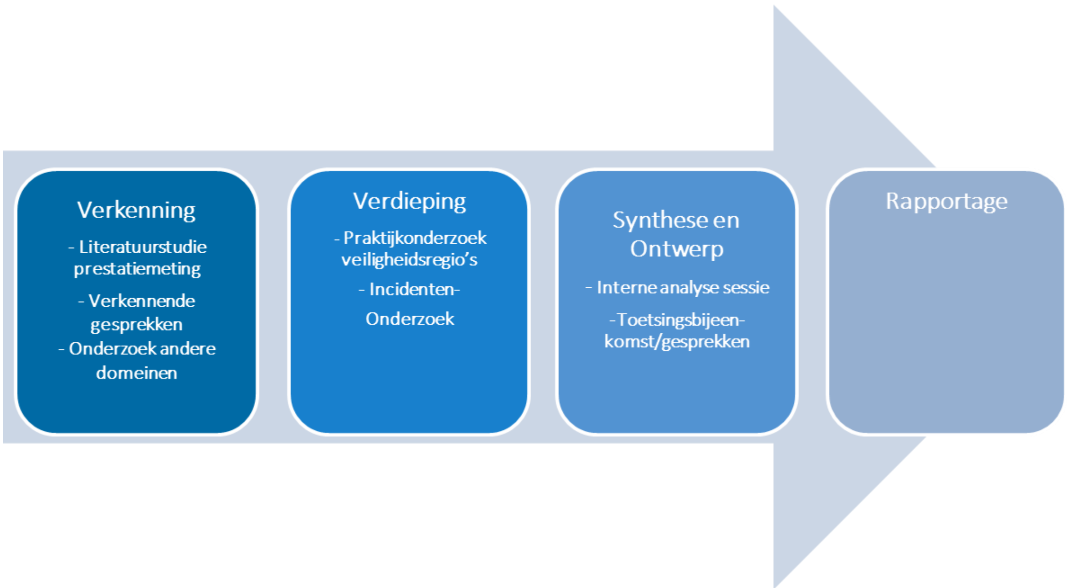
In het onderzoek staan zowel bestuurlijke als operationele prestaties van de veiligheidsregio centraal. Dit leidt tot een selectie van gesprekspartners waarin beide perspectieven vertegenwoordigd zijn.

Ons is verder opgevallen dat het begrip ‘indicator’ niet bij iedereen dezelfde associatie oproept. Een ‘indicator’ wordt meestal gezien als een kwantificeerbare operationalisering van dienstverlening – er is een product, de productie kan je tellen, zodat je tot een kwantitatieve score kunt komen. Er zijn ook respondenten die spreken over indicatoren, ook als deze geen kwantificeerbare operationalisering toestaan – er kan ook sprake zijn van een indicator, die alleen een kwalitatieve operationalisering toestaat en dus niet telbaar is. Wij spreken in het vervolg van dit rapport over een ‘indicator’ in de betekenis van kwantificeerbare operationalisering. In de gevallen dat gesprekspartners refereerden aan een meer kwalitatieve operationalisering van dienstverlening, dan hebben wij dit expliciet vermeld.

Het onderzoek richt zich op de mogelijkheid en wenselijkheid om te sturen op indicatoren die betrekking hebben op de *output* en *outcome.* Indicatoren op *output* niveau hebben betrekking op directe resultaten van de veiligheidsregio’s. Directe resultaten zijn bijvoorbeeld opkomsttijden van de gealarmeerde sleutelfunctionarissen. Indicatoren op *outcome* niveau richten zich vooral op de vraag of inspanningen bijdragen aan het uiteindelijke doel van een veiligheidsregio, bijvoorbeeld een effectieve rampenbestrijding en crisisbeheersing (zie voor definities: De Bruijn, 2007).

1.4. Onderzoeksaanpak

Figuur 1 bevat een weergave op hoofdlijnen van de aanpak van het onderzoek.



Figuur 1. Fasen van het onderzoek.

Toelichting op het onderzoeksproces:

- **Fase Verkenning.**
    - *Verkenning prestatiemeting in veiligheidsregio’s.* In deze fase is een literatuurstudie uitgevoerd naar de theorie van prestatiemeting en is een aantal verkennende gesprekken gevoerd met betrokkenen uit de wereld van de veiligheidsregio. Een overzicht van gesprekspartners in de verkennende fase is opgenomen in de bijlage.

    - *Verkenning prestatiemeting in andere sectoren/domeinen.* In deze fase is een groslijst opgesteld met mogelijke sectoren en domeinen waaruit lessen getrokken kunnen worden. In overleg met de begeleidingscommissie zijn vijf sectoren geselecteerd. Voor elk van de sectoren is een documentstudie uitgevoerd.
- **Fase Verdieping.**
    - *Praktijkonderzoek in veiligheidsregio's.* Op basis van de bevindingen uit de verkennende fase is in overleg met de begeleidingscommissie een selectie gemaakt van te onderzoeken regio's en is een protocol opgesteld voor de gesprekken in de verdiepende fase van het onderzoek. In het protocol zijn de typen gesprekspartners beschreven, de gespreksonderwerpen en is het doel van het gesprek nader omschreven. Er zijn per regio gesprekken gevoerd op drie niveaus: bestuurlijk, directie en operationeel. In totaal zijn negen (groeps)gesprekken gevoerd. Een overzicht van gesprekspartners in de verdiepende fase is opgenomen in de bijlage.
    - *Incidentenonderzoek.* In deze fase zijn uitgevoerde incidentenonderzoeken bestudeerd. De selectie van incidentenonderzoek is in overleg met de begeleidingscommissie vastgesteld.
- **Fase Synthese en Ontwerp.**
    - *Interne analysesessie.* In de vierde fase van het onderzoek zijn de theoretische en praktische inzichten uit de verkenning en verdieping samengebracht en door het onderzoeksteam geanalyseerd.
    - *Toetsingsbijeenkomst vertegenwoordiging veld.* Nadat we beelden hebben verzameld uit de drie geselecteerde regio's, organiseerden we een bijeenkomst met afgevaardigden uit overige regio's. Tijdens de bijeenkomst is getoetst of de afgevaardigden van andere regio's de percepties uit de drie geselecteerde regio's herkennen. Voor een overzicht van gesprekspartners in de toetsingsbijeenkomsten zie bijlage.
    - *Toetsende gesprekken met Inspectie VenJ en NCTV.* Tijdens deze gesprekken is getoetst of de betrokkenen de beelden uit het onderzoek herkennen. Voor een overzicht van gesprekspartners in deze gesprekken zie bijlage.
- **Fase Eindrapportage.** In deze fase zijn de bevindingen van het onderzoek opgesteld in een conceptrapportage die is voorgelegd aan de begeleidingscommissie. De reacties op de conceptrapportage zijn verwerkt in het eindrapport.

1.5. Begeleidingscommissie

Het onderzoek is uitgevoerd door de onderzoekers van de TU Delft[6] en KWINK groep en vond plaats in de periode van september 2014 tot juli 2015.

Dit onderzoek is begeleid door een begeleidingscommissie, die meerdere keren met de onderzoekers is samengekomen. De begeleidingscommissie heeft toezicht gehouden op de uitvoering van het onderzoek, heeft opgetreden als informatiebron en heeft de conceptrapportage en verschillende tussenproducten getoetst. De commissie bestaat uit de volgende leden:

- De heer prof. dr. A.F.A Korsten (voorzitter van de begeleidingscommissie)
- De heer drs. J.H.H. Mans
- De heer prof. dr. R. Torenvlied (Universiteit Twente)
- Mevrouw drs. H.I.J. Slock (NCTV)

[6] Faculteit der Techniek, Bestuur en Management.

- De heer drs. A.W. Dorst (Inspectie VenJ)
- De heer drs. T.L. van Mullekom (WODC)

1.6. Leeswijzer

Dit rapport vervolgt met een hoofdstuk (2) waarin inzichten uit de literatuur over prestatiemeting in de publieke sector aan bod komen. We zetten uiteen wat er bekend is over de functies en effecten van prestatiemeting en de omstandigheden waaronder die effecten positief of pervers zijn.

Hoofdstuk 3 gaat in op prestatiemeting in de praktijk. De ervaringen met prestatiemeting in andere sectoren en in het veiligheidsdomein worden beschreven.

In hoofdstuk 4 worden op basis van de bevindingen uit de hoofdstukken 2 en 3 ontwerpeisen geformuleerd. De gedachte hierachter is dat een stelsel van indicatoren ten minste aan deze eisen zal moeten voldoen, wil het betekenisvol en kansrijk zijn.

De beelden van gesprekspartners uit de veiligheidsregio's over de wenselijkheid van een stelsel van prestatie-indicatoren worden beschreven in hoofdstuk 5.

In hoofdstuk 6 passen we de ontwerpeisen uit hoofdstuk 4 toe op de beelden van gesprekspartners uit de veiligheidsregio's. Op basis daarvan trekken we conclusies.

We sluiten het rapport af met een hoofdstuk reflectie (7), waarin we reflecteren op de vraag hoe (beter) inzicht kan worden verkregen in het presterend vermogen van veiligheidsregio's.

In de bijlagen zijn een overzicht van de geraadpleegde bronnen en een lijst met gesprekspartners opgenomen.

2. Prestatiemeting in literatuur

2.1. Inleiding

In dit hoofdstuk zal een overzicht worden gegeven van de literatuur over prestatiemeting in de publieke sector. We beschrijven mogelijke functies en effecten van prestatiemeting en identificeren condities waaronder prestatiemeting tot positieve of perverse effecten leidt. Het hoofdstuk sluit af met bevindingen die van belang zijn bij het beantwoorden van de vraag in hoeverre prestatie-indicatoren in theorie een geschikt instrument zijn om een beeld te vormen van het presterend vermogen van veiligheidsregio’s.

2.2. Twee verhaallijnen

In de wetenschappelijke literatuur zijn twee verhaallijnen over prestatiemeting in de publieke sector sterk vertegenwoordigd.

In de eerste lijn wordt het belang van prestatiemeting sterk benadrukt. De eenvoudige constatering is dat een overheid geld beschikbaar stelt voor een bepaalde activiteit en dus inzicht wil – in de belangrijkste doelen, in de streefcijfers, in de gerealiseerde prestaties. Veel van die prestaties zijn in cijfers uit te drukken - en die cijfers moeten dus gegenereerd worden. Scholen mogen bevraagd worden op aantallen geslaagde leerlingen, rechters op aantallen gewezen vonnissen, de politie op het percentage opgeloste misdrijven, de NS op vertragingen, enzovoort. Weinigen ontkennen dat die cijfers er toe doen en vaak wordt beredeneerd dat aandacht voor die cijfers tot verbetering van prestaties leidt (Osborne en Gaebler, 1992). 'What gets measured, gets done' is het bijpassende adagium (zie ook Behn, 2003; Wilson et al., 2006). Het is op z'n minst een interessant gegeven dat rechtbank A veel minder vonnissen wijst dan rechtbank B. Allerlei verklaringen kunnen dat verschil legitimeren. Maar soms kan blijken dat rechtbank B door een betere organisatie of door andere werkwijzen meer vonnissen kan wijzen – en dat dat een goede zaak is. Of neem het bekende voorbeeld van de afdeling Hartchirurgie van het Academisch Ziekenhuis Nijmegen. Een aantal jaren gelden bleek dat een hoog mortaliteitscijfer te hebben. De eerste reactie van het ziekenhuis was defensief van aard: dat hoge cijfer zou niets zeggen. Na verder onderzoek bleek het cijfer wel iets te zeggen – het werd veroorzaakt door slechte samenwerking tussen de artsen en dus door slechte dienstverlening. Dankzij het prestatiecijfer kwam dit falen van professionals aan het licht (Externe Onderzoekscommissie, 2006). Prestatiemeting levert belangrijke inzichten op en moet dus – en wie zich, in deze eerste verhaallijn, in het publieke debat kritisch uitlaat over de betekenis van prestatiemeting, wordt er zomaar van verdacht bang te zijn om afgerekend te worden.

In de tweede verhaallijn wordt het idee ontkracht dat prestaties in het publieke domein kunnen worden gemeten of tot gewenste resultaten leiden. Volgens critici komt het idee van prestatiemeting voort uit een neoliberale ideologie waarin de overheid als een bedrijf wordt gezien en overheidsprestaties dus gemeten kunnen worden. Inherent aan overheidsprestaties is dit dat vaak niet kan. De kwaliteit van onderwijs, rechtspraak, politie, openbaar vervoer of zorg laat zich niet in een of enkele cijfers vangen. Veel van die activiteiten zijn relationeel – zorg, onderwijs, rechtspraak gaan over relaties, over bejegening. Relaties vang je niet in een cijfer. Bij veel prestaties gaat het bovendien om kwaliteit – niet om kwantiteit. De kwaliteit van de lessen van een docent, de juridische kwaliteit van een vonnis, de professionaliteit van een agent – het is niet in cijfers uit te drukken.

Zodra wel wordt gepoogd om kwaliteit in cijfers uit te drukken, wordt die prestaties onrecht aan gedaan. Daarnaast vrezen critici voor het ontstaan van 'rendementsdenken'. Zie de kritiek op het universitaire onderwijs, waar goede rendementen belangrijker zijn dan goed onderwijs. Of zie de kritiek op het wetenschappelijke onderzoek, waar aantallen artikelen, aantallen citaties en omvang van de gerealiseerde financiering belangrijker zijn dan de inhoud van het onderzoek.

De opkomst van prestatiemeting is bovendien ingebed in de opkomst van managers in professionele organisaties. Die organisaties raken door MBA-tools als prestatiemeting over-managed, hetgeen ten koste gaat van de professionaliteit van dit soort organisaties en van de noodzakelijke professionele autonomie. Er kunnen vraagtekens worden gezet bij deze wijze van management en sturing. En het meer fundamentele probleem is ideologisch van aard: de neoliberale ideologie die dit alles in werking heeft gezet, maakt een fundamentele denkfout. De overheid is geen bedrijf. De zorg en al die andere sectoren zijn geen markten (Van den Brink et al., 2005).

Uit het onderstaande literatuuroverzicht blijkt dat beide verhaallijnen een kern van waarheid bevatten en dat het van belang is om per situatie na te gaan of prestatiemeting bevorderlijk kan zijn voor overheidsprestaties.

2.3. De essentie: meetbaar, relevant en toerekenbaar

De centrale gedachte van prestatiemeting is dat een prestatie van een overheid kan worden gekwantificeerd. De vaak ingewikkelde prestatie van, bijvoorbeeld, een rechter kan gereduceerd worden tot of geoperationaliseerd worden in een meetbare eenheid: het aantal door de rechter gewezen vonnissen, eventueel onderverdeeld naar verschillende typen vonnissen. Natuurlijk zeggen aantallen vonnissen niet alles, maar het zegt wel iets. Als de output per rechter of per rechtbank bekend is, kunnen vergelijkingen gemaakt worden, streefcijfers geformuleerd worden en met behulp van de cijfers kunnen rechtbanken kritisch worden bevraagd. Kortom, prestatiemeting maakt sturing mogelijk (De Bruijn, 2007).

Bij het gebruik van prestatiemeting spelen drie aspecten een belangrijke rol: meetbaarheid, relevantie en toerekenbaarheid.[7]

Meetbaarheid

Meetbaarheid kan een probleem zijn. Ten eerste kan onderscheid gemaakt worden tussen output en outcome (De Bruijn, 2007). Bij output gaat het om de directe en meest zichtbare resultaten. In het geval van de rechter is dat het vonnis. Bij outcome gaat het om de beoogde effecten. Een beoogd effect van rechtspraak kan bijvoorbeeld conflictbeslechting zijn of een goed functionerende rechtstaat. In het rijtje vonnis - conflictbeslechting - goed functionerende rechtstaat, wordt het vonnis 'output' genoemd, een goede functionerende rechtstaat de 'outcome' en conflictbeslechting een 'intermediair effect', dat zich tussen output en outcome bevindt.

Het kan voorkomen dat een rechter een heel hoge output heeft (veel vonnissen), maar dat het ene na het andere vonnis in hoger beroep wordt vernietigd. Dan scoort deze rechter slecht op het intermediaire effect. En

---

[7] Deze aspecten zijn afgeleid van de theorie over het SMART-principe. SMART staat voor 1) specifiek 2) meetbaar 3) aanwijsbaar 4) relevant 5) tijdsgebonden. In dit rapport wordt over drie van de vijf aspecten verder uitgeweid omdat ze op gespannen voet met elkaar kunnen staan. De andere twee aspecten zijn voor dit onderzoek minder relevant omdat ze geen spanning vertonen met de andere aspecten.

welke bijdrage levert deze rechter, met de vele vonnissen die worden vernietigd, aan een goed functionerende rechtsstaat?

Het probleem met output en outcome is dat het uiteindelijk gaat om de outcome, maar dat die lastig te meten is en bovendien minder goed toerekenbaar is. Dus verlaten we ons vaak op output, die goed meetbaar is en waarvan de toerekenbaarheid zich ook vaak goed laat vaststellen.

Naarmate we minder weten over outcome of de relatie tussen output en outcome minder eenduidig is, neemt de betekenis van prestatiemeting af en wordt het gebruik van prestatiemeting meer problematisch.

Ten tweede speelt bij meetbaarheid het multi-waarden karakter van prestaties een rol. Prestaties in het publieke domein zijn altijd multi-waarden prestaties: ze moeten aan verschillende, vaak onderling conflicterende criteria voldoen. Een docent die een vak aanbiedt, moet zorgen dat het lesmateriaal up to date is, dat hij begrijpelijk is voor studenten, dat de stof voldoende uitdagend is, dat de les didactisch verantwoord is, dat het vak inhoudelijk aan professionele standaarden voldoet en dat het alles bij elkaar studeerbaar blijft voor studenten. Slechts een enkele waarde is kwantificeerbaar – in dit voorbeeld is het bijvoorbeeld denkbaar dat studeerbaarheid zich laat kwantificeren (in doorlooptijden, bijvoorbeeld).

Bovendien conflicteren de waarden – studeerbaarheid is belangrijk, maar uitdaging bijvoorbeeld ook en tussen die twee bestaat een spanning. Een eenzijdig accent op studeerbaarheid kan ten koste gaan van de uitdaging.

Hieruit volgt de conclusie dat naarmate minder aspecten van de prestatie kwantificeerbaar zijn en naarmate meer waarden conflicterend zijn, de betekenis van prestatiemeting afneemt of, anders geformuleerd, meer problematisch wordt.

Relevantie

Daarnaast moeten indicatoren relevant zijn. Als het al lukt om goed meetbare indicatoren te vinden, dan zijn die niet altijd relevant. Een voorbeeld: het aantal aangiftes is goed meetbaar en is een indicator voor de gerealiseerde veiligheid. Maar hoe relevant is die indicator? Betekenen veel aangiftes per definitie dat de wijk, gemeente of stad onveilig is? Of zijn de vele aangiftes het gevolg van een succesvol project waarin de procedure voor aangifte is vereenvoudigd? Zoals gezegd, het gaat om outcome, maar we verlaten ons vaak op output, omdat die nu eenmaal goed meetbaar en toerekenbaar is. Een aangrijpingspunt voor management zo lijkt het. Maar die output is echter niet altijd relevant – die raakt niet altijd aan de kern van het presteren van een organisatie. Output is niet interessant, als niets bekend is over het uiteindelijk beoogde effect – als de relatie tussen output en outcome niet duidelijk is vast te stellen.

Toerekenbaarheid

Tot slot moet indicatoren toerekenbaar zijn. Toerekenbaarheid heeft betrekking op de vraag of de prestatie die wordt gemeten, toerekenbaar is aan degene die de prestatie op papier levert. Stel dat een rechter een hoge productie heeft en stel dat we dat positief waarderen. Is die dan geheel toerekenbaar aan de betreffende rechter? Dat hoeft niet zo te zijn. Misschien beschikt deze rechter toevallig over heel goede niet-rechterlijke ondersteuning in de vorm van gerechtssecretarissen. Misschien zijn de advocaten in de regio meer meegaand dan elders in het land. Misschien doet het Openbaar Ministerie goed haar werk: het levert dossiers op tijd en volledig aan, waardoor de zittingen goed te plannen zijn. Een rechter bij een andere rechtbank heeft wellicht minder goede ondersteuning, meer assertieve advocaten en een Openbaar Ministerie dat minder goed plant.

Toerekenbaarheid is vooral een issue als een prestatie het resultaat is van samenwerking tussen verschillende spelers. In ketens of in netwerken is een prestatie vaak lastig eenduidig toe te rekenen aan een of enkele spelers.

2.4. Effecten van prestatiemeting

De literatuur geeft inzicht in zowel positieve als negatieve of 'perverse' effecten van prestatiemeting.

Prestatiemeting kan bevorderlijk zijn voor de productiviteit van een organisatie. Ze kan leiden tot meer publicaties (wetenschap), een hoger oplossingspercentage (politie), meer medische handelingen (zorg), meer onderhoud voor eenzelfde prijs (infrastructuur) (De Bruijn, 2008; Eshuis, 2008; Schoenmaker, 2012). Vaak wordt productie als 'slechts' kwantiteit gezien, maar toename van productie kan ook leiden tot hogere kwaliteit. Een wetenschapper die meer publiceert, moet nog altijd door de peer review van collega-wetenschappers komen. Kwantiteit kan tot meer kwaliteit leiden als wordt gepubliceerd in verschillende tijdschriften met hoge kwaliteitsstandaarden. Een arts heeft zijn professionele waarden en zal een standaard hebben waaraan zijn medische ingrepen moeten voldoen - meer kwantiteit hoeft niet te betekenen dat er minder kwaliteit wordt geleverd.

Toch kunnen zich ook perverse effecten voordoen. De literatuur rapporteert veelvuldig over het verschijnsel van 'gaming the numbers' (Goddard, 2000; Schoenmaker, 2012; Kerpershoek, 2015). Op papier wordt de productie verhoogd, maar het gaat slechts om productie op papier. In de werkelijkheid neemt de productie niet toe. Hoe krijg je de wachttijden op een EHBO-afdeling omlaag? Door patiënten buiten de wachtkamer te laten wachten, dan hoeven ze niet geregistreerd te worden (Goddard, 2000). Hoe krijg je Cito-toetsscores omhoog? Door zwakke leerlingen zich te laten ziek melden (Visser, 2003). Hoe produceer je meer vonnissen? Knip een vonnis op en wijs vaker een tussenvonnis. Prestatiemeting kan dus leiden tot een toename van waardevolle productie, maar ook tot productie op papier - en het is schadelijk als daar de verkeerde conclusies aan verbonden worden. Bijvoorbeeld als het beeld ontstaat dat de wachttijden zijn afgenomen of prestaties van leerlingen verbeterd zijn.

Er is hier nog een aspect dat aparte vermelding behoeft. Prestatiemeting is een prikkel voor productie, maar soms is toename van productie onwenselijk. Denk aan de zorg, waarin juist alles in het werk wordt gesteld om de vraag naar zorg niet te veel te laten stijgen (Kerpershoek, 2015).

Prestatiemeting kan tot innovaties leiden. Wie bijvoorbeeld beter moet presteren, zal nadenken hoe dat kan – welke vernieuwing mogelijk is. Doorlooptijden van promoties zijn te lang? Of rechtszaken nemen te veel tijd in beslag? Ongetwijfeld zal inzicht in deze cijfers leiden tot ideeën over organisatorische of inhoudelijke innovaties om de doorlooptijden te verkorten. Ook kan prestatiemeting tot een gesprek leiden over de onderliggende professionele waarden. Wat is eigenlijk een goed vonnis? Wat is eigenlijk een goede dissertatie (Pen, 2009)?

Ook hier is er een keerzijde. Prestatiemeting kan innovaties ook afremmen en tot risicomijdend gedrag leiden. Wie tot 'productie' wordt gedwongen, zal geneigd zijn om het bestaande te reproduceren. De wetenschapper die een publicatiedruk voelt, kan er voor kiezen om bestaande inzichten eindeloos te repliceren. Hij voelt geen prikkel om aan een nieuwe onderzoeksagenda te werken, omdat het aantal publicaties dan tijdelijk afneemt (Pen, 2009). Of hij spant zich vooral in om de gewenste productie te realiseren, niet meer en niet minder, zodat hij een ‘vinkje’ krijgt van zijn bovengeschikte (zie ook Ridderstrale en Nordstrom, 2004).

Achter prestatiemeting gaat een sympathiek idee schuil. Een bestuurder of een professional doet iets dat in de regel ingewikkeld is. De wijze waarop hij zijn producten of diensten vervaardigt, daar gaat de professional of bestuurder over. Die heeft recht op een zekere autonomie. In ruil daarvoor kan de professional of bestuurder worden aangesproken op zijn prestaties. Het klassieke voorbeeld is de hoogleraar. Die wordt geacht goed onderzoek te doen. De manier waarop doet niet ter zake, maar hij kan wel worden aangesproken op zijn output in de vorm van wetenschappelijke publicaties.

Een voordeel van prestatiemeting kan derhalve zijn dat het tot de-bureaucratisering leidt. Interne proces-voorschriften (voorschriften die aangeven hoe werkzaamheden uitgevoerd moeten worden) doen er niet meer toe – het gaat om het resultaat (De Bruijn, 2007).

Ook dit voordeel heeft een spiegelbeeld. Systemen van prestatiemeting kunnen tot nieuwe bureaucratie leiden. Allereerst hebben systemen van prestatiemeting de neiging om uit te dijen. De verklaring daarvoor is eenvoudig. Cijfers geven een beperkt beeld van de werkelijkheid. Wie een beter beeld van de werkelijkheid wil hebben, vraagt dus al snel om meer cijfers. Die vraag kan komen van de bovengeschikte. Wie er lucht van heeft gekregen dat scholen via ziekmeldingen Cito-toetsscores oppoetsen, wil wellicht een extra cijfers zien: het aantal ziekmeldingen. Wie als directeur van een school zich onheus bejegend voelt als zijn Cito-scores worden vergeleken met die van anderen, wil misschien ook wel een extra cijfer zien – bijvoorbeeld het opleidingsniveau van de ouders. Daarnaast leidt het 'gaming the numbers' soms tot een kat en muisspel. De muis speelt met de cijfers, de kat neemt allerlei maatregelen om gaten in het systeem te dichten: scherpere definities, extra controles, extra cijfers, et cetera. Hoe meer een systeem uitdijt, hoe bureaucratischer het wordt (Bouckaert en Halligan, 2007).

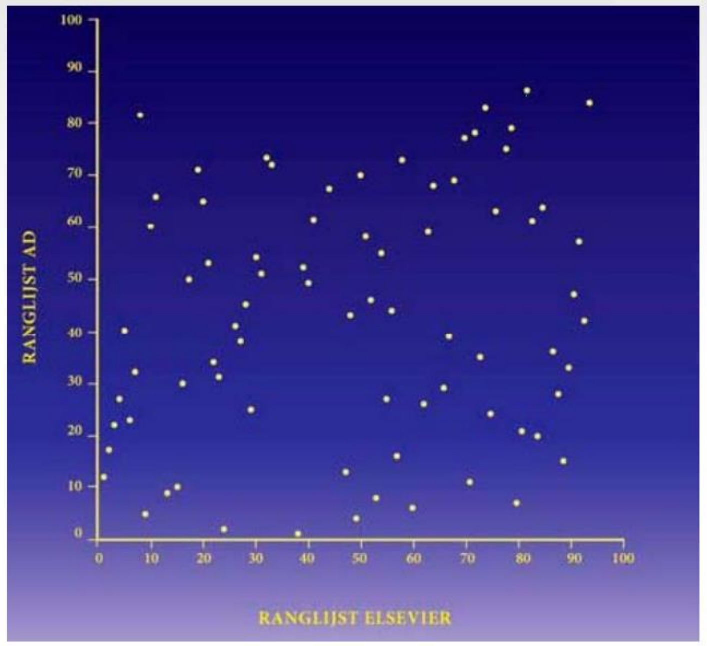
Prestatiemeting kan op nog een tweede manier tot bureaucratie leiden. Het basisidee is dat verantwoording afgelegd wordt over prestaties, maar dat de professional vrij is in de manier waarop hij producten en diensten tot stand brengt. In de taal van prestatiemeting: er wordt verantwoording afgelegd over output en/of outcome. Over het proces waarlangs die tot stand komt, hoeft geen verantwoording afgelegd te worden. Als outcome moeilijk te bepalen is, kunnen output-indicatoren ontwikkeld worden. Voor bepaalde activiteiten is het echter ook lastig om output-indicatoren te ontwikkelen. Dan bestaat het risico dat er een toevlucht wordt genomen tot proces-indicatoren (Schoenmaker, 2012). Die zeggen echter weinig of niets over output en outcome – waardoor een systeem van prestatiemeting juist bureaucratie brengt in plaats van ze tegengaat.

Ten slotte is prestatiemeting een krachtige manier van verantwoorden. Een politiekorps dat in cijfers kan laten zien dat de criminaliteit daalt, heeft een krachtiger verhaal dan een politiekorps dat alleen kwalitatieve verhalen levert. Wanneer een universiteit een groot aantal publicaties per wetenschapper heeft, zegt dat iets over de kwaliteit van die universiteit. Een ziekenhuis waarbij hoge doorligcijfers worden geconstateerd, overtuigt toch het meest als het die doorligcijfers omlaag weet te brengen.

Tegelijk kunnen cijfers ook tot een schijnverantwoording leiden. Het cijfer geeft immers maar een beperkt beeld van een prestatie. Denk hier aan middelbare scholen die kunnen worden gerankt op examenresultaten. Dat zegt iets, maar zegt ook heel veel niet. Ouders en leerlingen die zich bij de keuze voor een school laten leiden door die cijfers, zien over het hoofd dat andere aspecten ook belangrijk zijn bij de keuze voor een school.

Bovendien hangt het er vaak maar vanaf welke indicatoren worden gebruikt. Het plaatje hieronder laat een 'sterrenhemel' van ziekenhuizen zien. Op de ene as zijn ze van slecht naar goed gerankt, volgens criteria die het AD hanteert. Op de andere as volgens criteria van Elsevier. Blijkbaar zijn er ziekenhuizen die volgens de een tot de top behoren en volgens de ander slecht zijn – en andersom. Het laat de relatieve betekenis zien van prestatiemeting in verantwoordingsprocessen (Maarse en Van Velden, 2004). Schijnverantwoording ontstaat ook omdat soms geldt dat gewenste prestaties niet altijd genormeerd zijn. Wat is 'goed' als het om presteren gaat? In het eerder genoemde onderzoek naar de afdeling Hartchirurgie in Nijmegen beriep het ziekenhuis zich er op dat ze het prima deed, en boven het Europese gemiddelde scoorde. De vraag is of dat 'goed' is, als in de Europese score ook lidstaten worden meegenomen met een slecht ontwikkelde gezondheidszorg.

Concluderend, prestatiemeting kan leiden tot zowel positieve als perverse effecten.



bron: H. Maarse en M. van Velden (2004), Het ene beste ziekenhuis is het andere niet, in: *Medisch Contact*, nr. 44, pp 17-22

2.5. Wanneer positief, wanneer pervers?

In de literatuur worden condities geïdentificeerd waaronder prestatiemeting tot positieve of perverse effecten leidt. Wanneer domineren de positieve effecten en wanneer de perverse effecten? Er zijn drie antwoorden op die vraag.

Het eerste antwoord: dit hangt af van de impact van een systeem. Is die hoog of laag? Een hoge impact betekent dat het wel of niet behalen van een productiecijfer grote gevolgen heeft. Daarbij kan worden gedacht aan de volgende gevolgen:

- Financiële gevolgen: een budget of beloning is sterk afhankelijk van een gerealiseerde prestatie in cijfers.
- *Naming and shaming*: de prestaties worden publiek gemaakt – bijvoorbeeld in de vorm van een ranking – en dit kan van invloed zijn op de reputatie van een organisatie of het keuzegedrag van cliënten.
- Politieke, bestuurlijke of managerial aandacht: wie de gewenste prestaties niet haalt, komt onder een politiek, bestuurlijk of managerial vergrootglas te liggen – moet zich extra verantwoorden, krijgt te maken met verscherpt toezicht, moet verbeterplannen maken.

Het mechanisme dat high impact systemen oproepen, is eenvoudig. De impact is hoog, dus ontstaat er een prikkel voor perverse effecten (Schoenmaker, 2012; Pen, 2009; Kerpershoek, 2015). Het gaat steeds om gepercipieerde impact (Pen, 2009). Hoewel de impact van een systeem feitelijk niet erg hoog hoeft te zijn, kan die toch als hoog worden gepercipieerd. Door cijfers op papier iets te verbeteren, iets minder innovatie en door bureaucratie iets te verlagen, kunnen de cijfers er opeens veel beter uitzien.

Een tweede antwoord: dit hangt af van de perceptie van de professional – ervaart hij het systeem als oneerlijk of dom dan is de kans op pervertering groot. Uit onderzoek blijkt dat artsen systemen van prestatiemeting in

de zorg regelmatig doen perverteren. Ze registreren niet de verrichte handeling, maar een andere handeling – vaak één die meer geld oplevert. Dit wordt 'upcoding' genoemd. Dat lijkt volgens sommigen op fraude, maar artsen doen dit ook vanuit professionele overwegingen (Kerpershoek, 2015). De 'producten' in het systeem passen niet op de professionele werkelijkheid. Een arts verricht bijvoorbeeld een tijdrovende diagnose omdat hij een hersenaandoening vermoedt. Vervolgens blijkt er niets aan de hand en kan de patiënt met wat eenvoudige medicatie naar huis. Volgens het systeem van prestatiemeting is er dan niets aan de hand, dus kan de arts niets registreren. De arts ervaart dat als oneerlijk – hij heeft uitgebreid onderzoek gedaan. Of het systeem verbiedt een arts een medicijn voor te schrijven dat de patiënt nodig heeft. Voor de professional is daarmee sprake van een 'dom' systeem dat geen recht doet aan de professionele werkelijkheid. En daarmee voelt hij zich gelegitimeerd om het systeem te perverteren. Sterker, wie dat niet doet, doet volgens hem de kwaliteit van de dienstverlening tekort.

Hier kan een relatie worden gelegd met meetbaarheid, relevantie en toerekenbaarheid . Want wanneer worden systemen als oneerlijk of dom ervaren? Als de geleverde prestatie niet toerekenbaar is (oneerlijk), als ze moeilijk meetbaar is of als ze niet relevant is – en de gemeten prestatie dus 'dom' is, omdat ze weinig zegt over de werkelijke prestatie.

Tot slot is er een combinatie mogelijk van de twee voorgenoemde condities. Prestatiemeting heeft een hoge impact en doet geen recht aan de professionele werkelijkheid. Het zal duidelijk zijn dat er dan sterke prikkels zijn om prestatiemeting te doen perverteren. Je wordt met een oneerlijk en/of dom systeem geconfronteerd en daar ook nog hard op afgerekend? Dat voelt als een legitimering om het systeem te doen perverteren.

Het gevolg hiervan is vaak een uitdijend systeem – in een poging pervertering tegen te gaan. Het gevolg daarvan is dat prestatiemeting bureaucratiseert en verwordt tot afvinken.

Conclusie: perverse effecten zullen zich vooral manifesteren bij high impact systemen, die als unfair of dom worden ervaren. Problematische meetbaarheid, relevantie en toerekenbaarheid zullen het beeld van een oneerlijk en dom systeem voeden.

2.6. Functies van prestatiemeting

Hard afrekenen op prestatiecijfers is dus risicovol. Dat leidt tot de vraag wat de functies van prestatiemeting kunnen zijn. Vaak worden drie functies genoemd (Van Mil, 2008):

- Vergelijken en leren: met behulp van prestatiemetingcijfers kan inzicht worden verworven in een organisatie of kan een leerproces opgestart worden. De mogelijkheid om vergelijkingen te maken met andere organisaties kan ook bevorderlijk zijn voor leren.
- Transparantie en verantwoorden: prestatiemeting biedt een beeld (met het accent op 'een') van de prestaties van een organisatie. Het is daarmee een instrument om verantwoording af te leggen.
- Oordeelsvorming en afrekenen: prestatiemeting wordt gebruikt om tot oordelen over goed en fout presteren te komen, en om vervolgens af te rekenen (zie de passage over high impact systemen).

De kunst bij systemen van prestatiemeting is om tot een goede balans te komen. Wanneer het alleen om vergelijken en leren gaat, bestaat de kans dat er geen sturende werking van de systemen uitgaat – ze kunnen als ongevaarlijk en dus niet echt belangrijk beschouwd worden. 'Leren' is immers een ambigu begrip. Stel dat een actor slordig is, slecht presteert en echt iets moet veranderen. Als die actor alleen maar wordt uitgenodigd tot 'leren', kan die daarin een kans zien om steeds weer de dans te ontspringen. Wanneer het

oordeelsvorming en afrekenen is, is de kans groot dat er ook geen sturende werking van prestatiemeting uitgaat, omdat de systemen perverteren.

Hoe kan dat balanceren vorm krijgen? In de literatuur wordt daartoe een aantal suggesties gedaan (De Bruijn, 2007; Propper en Wilson, 2003; Pollitt, 2013).

- Biedt ruimte voor een zekere variëteit aan indicatoren. Dat voorkomt dat er een eenzijdige focus op een of enkele indicatoren kan ontstaan en biedt ruimte om de gecompliceerde werkelijkheid over het voetlicht te brengen.
- Erken dat niet alle typen prestaties in indicatoren kunnen worden gevangen – en biedt dus ook ruimte voor meer kwalitatieve beschouwingen.
- Baseer nooit een oordeel op uitsluitend de indicatoren – maar nodig de presterende partij altijd eerst uit om het verhaal achter de indicator te vertellen.
- Spreek onderling af welke indicatoren publiek zijn en welke alleen gebruikt worden in het verkeer tussen bijvoorbeeld de manager en de organisatie of persoon wiens prestaties worden gemeten.
- Wanneer na een gesprek over een indicator, de beoordelende partij van mening is dat er niet goed wordt gepresteerd, biedt de presterende partij dan eerst de kans om zelf tot verbetering te komen. Pas wanneer deze uitblijft, kan de indicator langzaamaan worden gebruikt voor afrekenen.

Deze suggesties zijn te vertalen in een aantal spel- of procesregels rond prestatiemeting. Die spel- of procesregels moeten de agent het vertrouwen geven dat prestatiemeting niet dom of oneerlijk wordt gebruikt of dat sprake is van een high impact systeem. Ze moeten de principaal het vertrouwen geven dat er met behulp van prestatiemeting echt wordt geleerd, verantwoording wordt afgelegd en, als het moet, dat de agent er ook op kan worden afgerekend.

2.7. Bevindingen

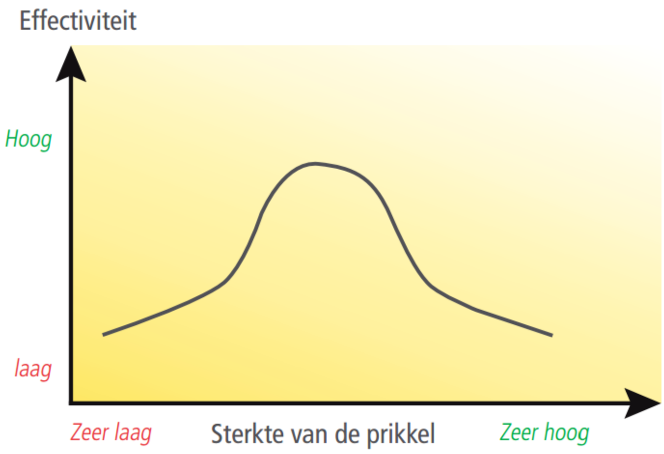
Dit overzicht van de literatuur levert een aantal inzichten op die we kunnen gebruiken bij onze zoektocht naar output- en outcome-indicatoren voor de veiligheidsregio's.

- De indicatoren dienen meetbaar te zijn (in volgorde van preferentie: outcome, output, proces) en mogen niet te veel concurrerende waarden kennen die niet meetbaar zijn.
- De indicatoren dienen zoveel mogelijk relevant te zijn.
- De indicatoren dienen zoveel mogelijk toerekenbaar te zijn.
- De best denkbare combinatie is meetbaarheid + relevantie + toerekenbaarheid + (gepercipieerde) hoge impact. Er gaat dan een sturende werking uit van het prestatiemeetsysteem en het risico van pervertering is beperkt *(kwadrant A in het schema hieronder).*
- Echter, deze situatie doet zich in publieke sectoren zelden voor. Meetbaarheid, relevantie en toerekenbaarheid zijn namelijk in veel gevallen lastig te verenigen.
- Als meetbaarheid, relevantie en toerekenbaarheid inderdaad problematisch zijn en de (gepercipieerde) impact van het prestatiemeetsysteem is hoog dan is het risico van pervertering groot en zal de sturende werking gering zijn. De slechtst denkbare combinatie is problematische meetbaarheid + problematische relevantie + problematische toerekenbaarheid + (gepercipieerde) hoge impact *(kwadrant B)*.
- De (gepercipieerde) impact mag dus niet te hoog zijn bij problematische meetbaarheid, relevantie en toerekenbaarheid– dat leidt tot perverse effecten.
- Het logische alternatief lijkt: reken niet hard af, maar richt prestatiemeting op 'leren'. Het risico is dan echter dat er weinig sturende werking van prestatiemeting uitgaat *(kwadrant C en D).*

	Meetbaarheid, relevantie en toerekenbaarheid niet problematisch	Meetbaarheid, relevantie en toerekenbaarheid problematisch
Lage impact	C. Weinig sturende werking, weinig prikkels voor pervertering	D. Weinig sturende werking, weinig prikkels voor pervertering
Hoge impact	A. Sturende werking, Weinig prikkels voor pervertering	B. Weinig sturende werking, sterke prikkels voor pervertering

Tabel 1 Factoren die sturende werking en pervertering beïnvloeden

- De effectiviteit van prestatiemeting is optimaal als een goede balans wordt gevonden in de sterkte van de prikkels doordat zowel aandacht is voor ‘leren’ als ‘verantwoorden’. Zie figuur hierna.



bron: Hans de Bruijn (2006), *Prestatiemeting in de publieke sector*, Den Haag: Lemma, pp.50-51

Figuur 2 Effectiviteit van prestatiemeting in relatie tot sterkte van prikkels

3. Prestatiemeting in praktijk

3.1. Inleiding

Nadat we in hoofdstuk 2 op grond van de literatuur een aantal lessen hebben getrokken over prestatiemeting, zal in dit hoofdstuk de praktijk centraal staan. Wat zijn de ervaringen met indicatoren voor het presterend vermogen op output- en/of outcomeniveau in verschillende sectoren? Wat kunnen we leren over de werking van prestatiemeetsystemen in de praktijk?

We selecteren vijf sectoren. Per sector worden – op basis van een vaste vragenset – de ervaringen met prestatiemeting geschetst, om zo een indruk te geven van de wijze waarop prestatiemeetsystemen in de praktijk worden toegepast. De informatie die wordt aangereikt, is niet uitputtend. Op basis van de beelden kunnen geen conclusies worden getrokken over de werking van de onderzochte prestatiemeetsystemen. Op basis van de analyse kan echter wel een aantal belangrijke noties en lessen worden gedestilleerd over het gebruik van indicatorensets voor het inzichtelijk maken van presterend vermogen.

Aan het einde van dit hoofdstuk beschrijven we ook de recente ervaringen in het veiligheidsdomein zelf met prestatiemeetsystemen. Daarbij zal de focus met name liggen op de inventarisatie van een variëteit aan projecten die elk op een eigen wijze een bijdrage leveren aan het in kaart brengen van presterend vermogen. Ook deze kennis biedt aangrijpingspunten om de centrale onderzoeksvraag te kunnen beantwoorden.

3.2. Selectie van sectoren

De keuze van de verschillende sectoren is met de begeleidingscommissie afgestemd. Binnen de selectie vallen verschillende type sectoren, zodat inzicht verkregen wordt in de werking van prestatiemeting in de praktijk onder verschillende omstandigheden. We beschrijven sectoren 1) binnen de veiligheidswereld om aansluiting te zoeken bij trends en ontwikkelingen in het veiligheidsveld, 2) sectoren waarin andere toezichthouders (dan de Inspectie VenJ) actief zijn, 3) andere publieke domeinen en 4) domeinen in het bedrijfsleven. Er is niet geselecteerd op een gewenste uitkomst, namelijk of de gekozen sectoren zich de ene keer goed en de andere keer minder goed lijken te lenen voor prestatiemeting.

Hierna volgt een korte beschrijving van de sectoren die in dit hoofdstuk worden uitgediept.

De sector *zorg* is onderzocht omdat het een sector is waarin een andere toezichthouder dan de Inspectie VenJ actief is. Lessen uit deze sector bieden aangrijpingspunten voor de opgave van de toezichthouder.

Door *universiteiten* wordt steeds vaker gebruik gemaakt van prestatiemeetsystemen als gevolg van de toenemende verzakelijking van de universitaire wereld. Onder druk van veranderende maatschappelijke opvattingen over de waarde van universiteiten neemt het belang om te kunnen rapporteren over prestaties toe.

De *politie* is een professionele organisatie waarmee de veiligheidsregio samenwerkt. Nadere analyse van prestatiemeting in de praktijk bij de politie biedt mogelijke aanknopingspunten voor prestatiemeting in het gehele veiligheidsdomein.

Onderhoud civiele infrastructuur is gekozen omdat in deze publieke sector al geruime tijd ervaring is opgedaan met prestatiemeting. Bovendien lijkt op het eerste gezicht in termen van complexiteit sprake van een relatief eenvoudige sturingsopgave. Interessant is bovendien dat in deze sector private organisaties in plaats van publieke organisaties worden aangestuurd.

Om de beeldvorming te complementeren, is een sterk bedrijfsmatige sector – te weten *off-shore gas* – gekozen. Op deze wijze kan in de praktijk worden nagegaan hoe de wijze waarop prestatiemeting in de publieke sector wordt uitgewerkt verschilt van de private sector.

3.3. Zorg

Beschrijving en analyse gebaseerd op onderzoek door Kerpershoek (2015).

Waarop heeft het prestatiesysteem betrekking?

Als gevolg van de ingrijpende veranderingen die sinds het begin van de 21e eeuw in de zorgsector zijn aangebracht, is prestatiemeting ook in deze sector doorgedrongen. Op vele terreinen worden via systemen prestaties van onderdelen van de zorgketen in kaart gebracht en gebruikt om de zorg te financieren en aan te sturen.

Toezichthouder Nederlandse Zorgautoriteit (NZa) heeft tot taak om de zorgmarkten te creëren (reguleren) en bewaken (toezicht) en speelt in de ontwikkeling van het DBC–stelsel (diagnose-behandel combinaties) een belangrijke rol. Door de ontwikkeling van een bepaalde markt regelmatig in beeld te brengen, geeft de NZa op een gestructureerde manier invulling aan haar regulerings- en toezichttaken.

Een belangrijke schakel in de zorgketen vormen de prestaties van ziekenhuizen. Het DBC-systeem vervult in het meten van prestaties in de zorg een essentiële rol. Er zijn DBC's voor zorg in ziekenhuizen, zorg die wordt aangeboden via de ggz, zorg in de geriatrische revalidatiezorg (grz) en forensische zorg. Het systeem is gebaseerd op het idee van *pay per performance* en werd ingevoerd in 2005.

Om de prestaties van ziekenhuizen in kaart te brengen, wordt per zorgdiscipline die in ziekenhuizen vertegenwoordigd is een groot aantal diagnoses onderscheiden. Aan ieder type diagnose wordt een behandeling gekoppeld. Hiermee ontstaan diagnose-behandel combinaties. Een DBC is dus een "zorgprestatie": "de zorg die zorgaanbieders leveren voor een specifieke zorgvraag."[8] Een ziekenhuis 'produceert' dus een bepaalde hoeveelheid DBC's. De DBC is de totale ziekenhuisbehandeling, het gehele behandeltraject, vanaf de diagnose van de specialist tot en met eventuele ziekenhuisbehandeling en bijbehorende nacontrole(s). Iedere DBC kent een prijs. Een arts die bij een patiënt dus DBC X toepast, krijgt de vooraf vastgestelde prijs die bij X hoort. Die prijs is gebaseerd op de gemiddelde kosten die met een DBC X zijn gemoeid.

Wat is het doel of wat zijn de doelen van het systeem?

Het DBC-systeem vervult drie functies: 1) DBC's vormen onderdeel van een financieringssysteem, 2) het DBC-systeem wordt gebruikt om marktwerking in de zorg te introduceren en 3) inzicht bieden in de zorg die ziekenhuizen leveren. DBC's maken het mogelijk de prestaties van ziekenhuizen te benchmarken. Vervolgens ontvangen maatschappen of ziekenhuizen een vergoeding op grond van de totale hoeveelheid DBC's die zij

---

[8] Zie: http://www.nza.nl/zorgonderwerpen/dossiers/dbc-dossier/veelgestelde-vragen/DBC-DOT/

hebben ‘geproduceerd’. Over deze prestaties wordt meer en meer onderhandeld tussen zorgverzekeraars en zorgaanbieders.

Wat is de typering van de prestaties die gemeten worden?

Er is een keten te construeren: de klacht van de patiënt – de diagnose van de arts (proces) – de behandeling die de arts voorschrijft (output)– de daadwerkelijke behandeling (intermediaire effect) – het uiteindelijke effect van die behandeling op de gezondheid van de patiënt (outcome).

Wanneer het om heel eenvoudige aandoeningen gaat (bijvoorbeeld, diagnose: hoofdpijn, behandeling: paracetamol of diagnose: gebroken been, behandeling: gips), is er vaak een ‘harde’ relatie tussen output en outcome – dankzij de paracetamol gaat de hoofdpijn over, het gips heelt het been.

Patiënten die zich in ziekenhuizen melden, hebben meestal meer complexe klachten – en dan kunnen de relaties tussen proces – output – outcome veel minder hard zijn en zijn er allerlei verstorende variabelen.

Wat is de impact van het systeem?

De impact van het systeem is erg hoog. De productie van DBC’s bepaalt het budget dat een ziekenhuis of maatschap ontvangt. Zelfs wanneer artsen in loondienst zijn (en hun inkomen dus niet direct afhankelijk is van het volume aan geproduceerde DBC’s), blijken zij toch alert te zijn op de vraag of de productie voldoende is – omdat de financiering van hun afdeling daarvan afhankelijk is of kan zijn.

Wat zijn de kenmerken van de groep die via het systeem wordt aangestuurd?

Het kenmerk dat opvalt, is variëteit. Er zijn grote verschillen tussen disciplines (oncologie of oogheelkunde), en tussen organisaties waarin de DBC’s plaatsvinden (bijvoorbeeld gewone en academische ziekenhuizen).

Welke effecten van het systeem zijn gevonden?

- Transparantie: er ontstaat inzicht in de prestaties van de zorg in Nederland en er is een min of meer geobjectiveerd systeem om middelen te alloceren.
- “De ziekenhuissector laat een stijgende trend in de kwaliteit en kostenverhouding zien […] Helaas worden momenteel voornamelijk procesuitkomsten gemeten. Die laten wel alle een verbetering zien in de loop der jaren.”[9] Of dankzij de DBC’s de prestaties in de zorg (in termen van outcome) zijn toegenomen en dus of de gezondheid van de patiënt is verbeterd, is (nog) niet bekend.
- Of dankzij de DBC’s de efficiëntie van de zorg, kosten van de zorg in ziekenhuizen, is afgenomen, is niet vastgesteld en daarover lopen de meningen uiteen. Het ministerie van VWS (2010, p. 33) rapporteert dat de “kwaliteit kostenverhouding van ziekenhuiszorg” een licht stijgende trend laat zien en dat de kosten van ziekenhuiszorg constant zijn gebleven (2,9% van het BBP in 2004 (€14,0 miljard) ten opzichte van 2008 (€17,4 miljard)) tegen verbeterende kwaliteitsmetingen. Maar een “causale relatie tussen het gevoerde beleid en kwaliteit en kostenontwikkeling” (p. 46) kan niet worden bewezen. Critici wijzen op de sterk gestegen beheerskosten die de invoering van de DBC-systematiek (van minder dan 10 tot circa 15 à 20 procent van de zorgkosten) en het probleem dat de all-in systematiek leidt tot “meervoudige interpreteerbaarheid”, die systeemeffecten zoals kostenstijgingen moeilijk achterhaalt (Kreis, 2012, p. 1891).
- Belangrijke perverse effecten van het DBC-systeem zoals extra uitgevoerde behandelingen, ‘upcoding’ – het moedwillig vergroten “van de casemix van een ziekenhuis met als doel de vergoeding te verhogen” – zijn ook bekend (zie bijvoorbeeld Hasaart, 2011; Kerpershoek, 2015). Het volume van deze effecten is

[9] Zie de rapportage Werken aan de zorg, http://www.werkenaandezorg.nl/rapport/kk-datarapport.pdf.

onbekend. Schattingen zijn dat deze ongewenste effecten optellen tot een miljard euro in de periode 2006-2008 (Hasaart, 2011) of 800 miljoen euro aan 'opvallende declaraties' in 2012 (NZa). Het betreft:
  - Over-registratie: er worden per patiënt meerdere DBC's geregistreerd, daar waar het systeem dat niet toestaat.
  - 'Creative diagnosis registration': er worden andere diagnoses geregistreerd dan feitelijk zijn verricht.
  - 'Creative treatment registration': er worden andere behandelingen geregistreerd dan feitelijk verstrekt.
  - 'Patient selection': er vindt patiëntselectie plaats om zo het volume van DBC's gunstig te beïnvloeden.
- Ook kan verklaard worden waarom professionals actief dan wel passief omgaan met perverse effecten. Een belangrijke legitimatie die medische professionals voor hun gedrag geven, is dat het systeem van DBC's niet past op de professionele werkelijkheid. Voorbeelden:
  - Een arts constateert dat bij een diagnose een bepaalde behandeling past – maar die DBC bestaat niet in het systeem.
  - Er is het probleem van de 'verdenkingsdiagnose'. Een arts verricht een diagnose omdat hij een ernstige aandoening vermoedt. Die veronderstelling is niet juist en vervolgens moet hij een DBC registeren, waarbij financiering hoort die niet in verhouding is met de kosten die zijn gemaakt. Er wordt bijvoorbeeld hersenvliesontsteking vermoed, een ruggeprik gedaan, vervolgens blijkt het loos alarm – er is slechts sprake van griep.
  - De kosten van die ruggeprik zitten niet in een DBC voor griep. Dus wordt een DBC hersenvliesontsteking opgevoerd.
  - Cosmetische behandelingen worden in principe niet vergoed, maar de grens tussen medisch en cosmetisch is erg dun. Artsen en zorgverzekeraars verschillen hierover vaak van mening.

Het is buitengewoon moeilijk om aan te geven waar het gedrag van artsen te rechtvaardigen valt omdat het systeem niet past op de professionele werkelijkheid – en waar het overgaat in fraude.

Welke lessen kunnen uit de sector worden getrokken?

- De enorme variëteit aan medische ziekenhuisbehandelingen laat zich niet makkelijk in een prestatiemeetsysteem vangen. Processen die leiden tot prestaties, zowel het stellen van een diagnose als een behandeling, zijn zeer complexe processen. De volledige keten in de medisch specialistische zorg van diagnose via behandeling en controle tot outcome is bovendien lang en wordt mede beïnvloed door allerlei andere interveniërende factoren.
- High impact systemen roepen perverse effecten op. Over de vraag of die te rechtvaardigen zijn, verschillen de meningen. Sommigen noemen het fraude, anderen zeggen dat dit onontkoombaar is – de zorg van de patiënt staat voorop, het systeem past niet op de professionele werkelijkheid.
- Wanneer een (1) high impact systeem, (2) niet past op de professionele werkelijkheid, is er voor professionals een rechtvaardiging om met dat systeem te spelen en perverse effecten te genereren.
- Het DBC-systeem was erg gedetailleerd (er waren meer dan 30.000 producten) en bood weinig ruimte voor flexibiliteit. Het DBC-systeem is inmiddels doorontwikkeld en in 2012 is DOT (DBC op weg naar transparantie) ingevoerd. DOT omvat een minder gedetailleerd stelsel van diagnose-behandelcombinaties (er zijn nog 'maar' ruim 4.000 producten). Het prestatiemeetsysteem is daarmee sterk vereenvoudigd. Er vindt bovendien een 'ongoing' proces van interactie plaats tussen medisch specialisten, bestuurders en verzekeraars om te evalueren waar het systeem niet functioneert en waar aanpassingen gemaakt moeten worden. Er is gepoogd de financiële consequenties voor medisch specialisten minder direct te maken, maar tegelijkertijd zijn sinds 2005 prikkels om efficiënter te werken geïntroduceerd en is stapsgewijs (10% in 2005, 20% in 2008, 34% in 2009 tot 100% in 2012) volledige prestatiebekostiging voor alle ziekenhuisbehandelingen ingevoerd. Maar of dit alles leidt tot hogere transparantie van het zorgstelsel in ziekenhuizen is nog maar de vraag (Hasaart, 2011).

3.4. Universiteiten

Beschrijving en analyse gebaseerd op onderzoek door Pen (2009).

Waarop heeft het prestatiemeetsysteem betrekking?

Het systeem van prestatiemeting heeft betrekking op de registratie van publicaties van academische onderzoekers (output). In een studie naar de doorwerking van prestatiemeting van wetenschappelijk onderzoek (Pen, 2009) worden systemen onderzocht waarin prestaties van onderzoekers worden bijgehouden. Het gaat om: technische natuurkunde, geneeskunde, economie, sociale wetenschappen. Opmerkelijk is dat de systemen veel typen prestaties buiten beschouwing laten – er wordt met een beperkte lijst tijdschriften gewerkt, het gaat om publicaties in die tijdschriften. Publicaties in tijdschriften die niet in het systeem zijn opgenomen, doen niet mee. Kortom, de assumptie is dat wetenschappelijke prestaties alleen in deze bladen kunnen worden geleverd, maar hieraan kan worden getwijfeld. Dit soort tijdschriften zijn bijvoorbeeld overwegend conservatief en monodisciplinair van aard. Er zijn twee soorten beloningen voor wie goed presteert in deze systemen. Soms leidt een hogere prestatie tot financiële beloning aan organisatie-eenheden. Een hogere prestatie kan daarnaast tot een beloning voor individuele onderzoekers leiden: zij krijgen meer onderzoektijd als ze goed presteren.

Wat is het doel of wat zijn de doelen van het systeem?

De doelen van het systeem zijn: 1) inzicht bieden in de prestaties van wetenschappers, 2) het stimuleren van wetenschappers tot beter presteren en 3) het alloceren van geld of tijd naar degenen die het beste presteren.

Wat is de typering van de prestaties die gemeten worden?

De kern van het prestatiemeetsysteem is een output-indicator: het resultaat van onderzoek wordt neergelegd in een publicatie. De outcome die het prestatiemeetsysteem inzichtelijk wil maken, is de mate waarin wetenschappelijk onderzoek van onderzoekers van de universiteit bijdraagt aan nieuwe wetenschappelijke inzichten en tot maatschappelijke vernieuwing. Dat is echter moeilijk te meten. In de systemen die zijn opgesteld, wordt daarom output gemeten. De uitkomsten van wetenschappelijk onderzoek worden geoperationaliseerd door het aantal publicaties per onderzoeker. In een aantal gevallen is daartoe een typologie van publicaties gemaakt (grofweg: van ‘zware’ tijdschriften tot ‘lichte’ tijdschriften). De relatie tussen een publicatie en wetenschappelijke of maatschappelijke vernieuwing is echter niet eenduidig. Er zijn veel intermediaire factoren. Een wetenschappelijke publicatie kan theoretisch ook outcome zijn – als ze een innovatie beschrijft, die iedereen meteen opvolgt, maar dergelijke artikelen zijn er nauwelijks.

Wat is de impact van het systeem?

De impact van het systeem is op papier erg hoog: strategische ontwikkeling en positionering van universiteiten en budgetten en onderzoekstijd van onderzoekers zijn afhankelijk van de geleverde output. Zie hieronder echter voor een nuance, die uit het empirisch onderzoek blijkt.

Wat zijn de kenmerken van de groep die via het systeem wordt aangestuurd?

Ook hier domineert het kenmerk ‘variëteit’. De populatie academische onderzoekers is heel divers. Niet alleen tussen de wetenschappelijke disciplines, maar ook binnen de vakgebieden bestaan grote verschillen in publicatietradities – ook waar het aantallen betreft.

Welke effecten van het systeem zijn gevonden?

- Een sterke gerichtheid op meer publiceren, terwijl vaak ook meer wordt gepubliceerd. Daarbij doet zich overigens een kip-ei vraag voor. Leidt prestatiemeting tot een publicatiecultuur of creëert een publicatiecultuur systemen van prestatiemeting?

- De productie van wetenschappelijke artikelen neemt toe dankzij het systeem, maar daaraan is een bovengrens – op enig moment is het systeem uitgewerkt en gaat er geen sturende werking meer van uit.
- Invloed op personeelsbeleid: sollicitanten die goed scoren op de indicatoren uit dit soort systemen, maken meer kans.
- Nauwelijks invloed op de kwaliteit van het onderzoek.
- Perverse effecten zijn, verrassenderwijs, moeilijk aan te tonen maar niet uit te sluiten. Taken die niet worden gemeten in het prestatiemeetsysteem, worden gewoon nog goed uitgevoerd – er vindt dus geen verschuiving plaats naar uitsluitend de goed meetbare taken.
- Het aantal perverse effecten valt mee – terwijl dit toch een high impact systeem is. De verklaring:
    - Onderzoekers hebben mogelijkheden om op andere manieren dan via het prestatiemeetsysteem geld (en dus ook tijd) te verwerven. Denk hierbij aan NWO-beurzen of Europese onderzoeksgelden. Vanuit het perspectief van de professional is het systeem dus minder high impact dan het op papier lijkt.
    - De bestuurders die met het systeem van prestatiesturing werken, komen zelf uit de professie. Zij kennen dus de beperkingen van het systeem en zijn gevoelig daarvoor, maar wijzen het gebruik op hoofdlijnen niet af. Professionals die hun vraagtekens zetten bij het systeem, vinden in de bestuurders een mede-professional, die die vraagtekens begrijpt.
    - Belangrijk is dat er naast het systeem van prestatiemeting ook andere systemen zijn die onderzoekers beoordelen. Denk hierbij aan een website als Research Gate, waarop impact-scores van individuele onderzoekers inzichtelijk worden gemaakt en bijgehouden wordt hoe vaak artikelen van onderzoekers worden geciteerd en gedownload, maar ook onderwijsbeoordelingen. Ook dat heeft een matigende werking. Het betekent namelijk dat aspecten die niet door het systeem van prestatiemeting worden ondervangen (zoals met name kwaliteit) toch aan de orde kunnen komen in een verantwoordingsrelatie.
    - Prestatiemeetsystemen kunnen in de handen van managing professionals beperkte effecten genereren. Zij dempen de potentiële werking van een systeem.

Welke lessen kunnen uit de sector worden getrokken?

- Prestatiemeetsystemen die wetenschappelijke onderzoeksoutput in kaart brengen, zijn rigide en niet in staat om met productieprocessen met een rijke variëteit, zoals academische onderzoeksoutput, om te gaan. Veel belangrijke aspecten van academisch onderzoek worden niet gemeten – de maatschappelijke betekenis, innovaties die niet meteen tot artikelen leiden, de impact van een artikel op de ontwikkeling van een discipline.
- Invoering van prestatiemeting leidt doorgaans tot een vergroting van de ‘productie’: de kwantitatieve output van onderzoek.
- Prestatiemeting leidde niet tot een significante verschuiving op instellingsniveau (bijvoorbeeld van onderzoeksbudgetten naar beter presterende onderzoekers).
- Hoewel de theorie veronderstelt dat een sterke en directe koppeling tussen prestaties en financiering tot perverse effecten zal leiden, werden deze op instellingsniveau niet aangetroffen. Op het niveau van individuele onderzoekers lijken zich wel perverse effecten voor te doen. Denk hierbij aan voorbeelden van fraude en/of plagiaat of het ‘uitmelken’ of recyclen van eerder onderzoek.
- De transparantie die een prestatiemeetsysteem zou moeten brengen, is beperkt omdat veel relevante aspecten niet worden gemeten.

3.5. Politie

Waarop heeft het prestatiesysteem betrekking?

Bij de Nederlandse politie vormen prestaties en afspraken over prestaties al geruime tijd een belangrijk element in de wijze van aansturing. Sinds 2003 is sprake van nationaal beleid inzake prestatiesturing (zie verder Dijkzeul et al., 2008). De prestatieafspraken die met prestatiesturing werden afgesproken, waren SMART, beperkt op hoofdlijnen en gericht op output en doelmatigheid. Er is dus al geruime tijd sprake van een koppeling tussen afspraken en prestaties.

Wat is het doel of wat zijn de doelen van het systeem?

Doel van prestatiesturing bij de politie is tweeledig. Aan de ene kant: “bijdragen aan het presterend vermogen van de politie”. Daarnaast heeft prestatiesturing tot doel de prestaties “meer zichtbaar” te maken (CapGemini, 2007, p. 12). Hoewel dit niet als doel is gespecificeerd, kon via benchmarking het prestatiemeetsysteem worden gebruikt om te leren en wordt in rapportages ook gesproken over afrekenen.

Wat is de typering van de prestaties die gemeten worden?

Prestatieafspraken omvatten niet alle werkvelden en niet het gehele politiewerk. In plaats daarvan zijn voor specifieke taken afspraken gemaakt in termen van te behalen doelen. Vanaf 2004 zou 15% van de doelstellingen voor 2006 moeten zijn gerealiseerd en in latere jaren was sprake van een oplopend percentage doelbereiking (Dijkzeul et al., 2008, p. 26). De systemen meten momenteel throughput en outputprestaties van de politie. Gegeven de complexiteit van de relatie tussen output en outcome en het belang van de contextuele duiding van de indicatoren kunnen geen outcome-effecten geclaimd worden op basis van de prestatie-indicatoren (zie hieronder).

Er was de nodige kritiek op de wijze van prestatiemeting en de effecten (Dijkzeul et al., 2008). Zo ontstond de nodige ophef over de indicator en prestatie-eis van aantal boetes resulterend uit staandehoudingen. Hoewel de huidige prestatiemeetsystematiek in staat lijkt om de kwantiteit van de prestaties van de politie in kaart te brengen, schiet zij tekort in het meten van de kwaliteit van deze prestaties. Bestaande modellen van productiviteits- en prestatiemeting blijken met name problematisch in termen van validiteit en betrouwbaarheid. Sinds de vorming van de nationale politie zijn onderzoeken gedaan om een model van presterend vermogen van de politie (Zouridis, 2011) en productiviteitsmeting (Zouridis et al., 2014) te ontwikkelen. Deze modellen zijn echter complex en dit gaat volgens diezelfde onderzoekers ten koste van de bruikbaarheid en herkenbaarheid. Bovendien blijkt zelfs dat in het geval van complexe modellen voor de interpretatie en betekenisgeving ten minste ook aandacht dient te bestaan voor de specifieke context waarbinnen prestaties plaatsvinden, de perceptie van betrokken stakeholders en overeenstemming over de validiteit van het model dat de prestaties kan isoleren ten opzichte van maatschappelijke trends en ontwikkelingen. Tot slot blijkt de meest recente onderzoeken dat ook sprake is van een fundamentele onduidelijkheid over wat onder de productiviteit van de politie dient te worden verstaan (Zouridis et al., 2014).

Wat is de impact van het systeem?

De impact van het systeem is groot. Korpsen kregen op basis van de prestaties budget uitgekeerd. Hoewel dit slechts een klein gedeelte van het totale budget van de politiekorpsen betrof, werd de impact van prestatiemeting door de korpsleidingen als groot ervaren. Verder werd prestatiemeting door politiekorpsen overwegend positief geëvalueerd (focus op kerntaken) als kon worden voldaan aan de randvoorwaarde dat korpsen onderling goed vergelijkbaar zouden zijn. (CapGemini, 2007). Daarnaast bestond bezorgdheid over de weging tussen kwalitatieve en kwantitatieve factoren en werd eveneens gewezen op de perverse effecten van de benchmarking.

Wat zijn de kenmerken van de doelgroep die via het systeem wordt aangestuurd?

Ook hier domineert het kenmerk ‘variëteit’. Denk aan het decentrale en sterk contextuele karakter van politieoptreden en het complexe karakter van hetgeen gemeten wordt. Hoewel eenduidigheid in het meten van prestaties ter hand is genomen en de vorming van de nationale politie de regionale variëteit in politieoptreden heeft gereduceerd, concludeert recent onderzoek dat dan niet zomaar sprake kan zijn van een eenduidige wijze van meten van politieproductiviteit (zie Zouridis, 2011; Zouridis et al., 2014).

Welke effecten van het systeem zijn gevonden?

- Prestatiesturing heeft geleid tot meer en betere sturingsinformatie voor korpsbeheerders. Hoewel sprake was van scepsis vooraf heeft invoering van prestatiemeting geleid tot meer transparantie en resultaatverantwoordelijkheid en zijn – in tegenstelling tot waarvoor werd gewaarschuwd in vooronderzoek – de eerste rapportages positief wat betreft het zich voordoen van perverse effecten (Dijkzeul et al., 2008; CapGemini, 2007). Zo leek sprake van een dempende werking vanuit professionals en hun professionele autonomie, die perverse effecten van prestatiesturing tegen zou kunnen gaan (CapGemini, 2007, p. 20). In latere publicaties (zie hieronder) worden perverse effecten echter wel benoemd.
- Door de opkomst van prestatiemeting en prestatieafspraken is het genereren, analyseren, opslaan en ontsluiten van sturingsinformatie “aanzienlijk” toegenomen (CapGemini, 2007).
  Zouridis (2011) constateert dat prestatiedefinities tussen belangrijke groepen stakeholders (politiepersoneel, bestuurders en onderzoekers) significant verschillen. Dat is geen effect van het systeem, maar wel een belangrijke uitkomst die is gebleken na invoering van het systeem
- Realisatiecijfers laten een verhoogde output zien op de prestatieafspraken, maar over de outcome konden geen uitspraken worden gedaan. Prestatiesturing heeft op het terrein waarover specifieke afspraken zijn gemaakt geleid tot kwantitatieve positieve effecten op het presterend vermogen van de politieorganisatie. Op basis van een evaluatie van het systeem werd geconstateerd dat realisatiecijfers een verhoogde output laten zien op de prestatieafspraken, maar over de outcome konden geen uitspraken worden gedaan. Prestatiesturing lijkt vooral in kwantitatieve zin positief bijgedragen te hebben aan het presterend vermogen van de politieorganisatie, voor ten minste de gemaakte afspraken (Jochoms, 2011).
- Recent onderzoek laat zich vanwege een beperkt beeld van de outcome kritisch uit over de meerwaarde van het systeem, waarbij wordt gewezen op problemen rond het meten van kwantitatieve prestaties en de verdringingseffecten van ogenschijnlijk simpele outputcriteria die kwaliteit of moeilijk te meten processen buiten beeld van managers plaatst (zie Zouridis, 2011; Zouridis et al., 2014).
- Recent onderzoek wijst op verdringingseffecten van prestatiemeting: “Deze verdringing wordt veroorzaakt door de behoefte aan een uniform ‘productiesysteem‘ en de complexe context van zowel criminaliteit als het politieoptreden daartegen. Het aggregeren van politieoptreden naar algemene ‘productienoemers‘ doet niet alleen onrecht aan de dagelijkse politiepraktijk, maar verstoort en vervormt deze ook. Sommige prestaties worden wel zichtbaar, andere niet” (Zouridis, 2011, p. 117).
- “Voor prestatiemeting bij de politie geldt in het bijzonder het gevaar van overrepresentatie van specifieke taken (opsporing), de kwantiteitsbias die door de outputmetingen in de strafrechtketen kan optreden (aantal aangeleverde verdachten, et cetera) en de problematische relatie tussen output en outcome en de activiteiten van de politie en het daaruit voortvloeiende gevaar van doel-middel verschuiving.” (Zouridis et al., 2014, p. 7).

Welke lessen kunnen uit de sector worden getrokken?

- Voor simpele politietaken ontbreken 'natuurlijke' prestatiedefinities en prestatiemaatstaven. Ook de wettelijke taken zijn voor het bepalen van de prestaties "beperkt bruikbaar".[10] "Het definiëren van prestaties van de politie is een normatieve exercitie, zelfs als de wettelijke taken als vertrekpunt worden genomen". (Zouridis, 2011, p. 114).
- In de praktijk veelgebruikte indicatoren voor prestaties zijn sterk omstreden in termen van betrouwbaarheid en validiteit.
- Invoering van prestatiemeting heeft bij de politie plaatsgevonden tegen een achtergrond van toenemende centralisatie zonder dat dit heeft geleid tot verbetering in prestaties in termen van outcome.
- Er kan een aantal randvoorwaarden voor prestatiemeting worden opgesteld (CapGemini, 2007, p. 3):
    - Er moet sprake zijn van voldoende dialoog om prestatieafspraken en realisatiecijfers in perspectief te plaatsen.
    - Een actieve bewaking van zingeving van politiewerk moet voorkomen dat prestatiesturing leidt tot handelen naar de letter en niet naar de geest van sturingsprikkels.
    - Er moet voldoende ruimte zijn voor het lokaal bestuur om het beleid te adopteren in het eigen lokaal veiligheidsbeleid.
    - De verschillende sturingssystematieken van partners in de keten moeten op elkaar worden afgestemd om aansluiting van afspraken tussen ketenpartners tot stand te brengen.

3.6. Onderhoud civiele infrastructuur

Beschrijving en analyse gebaseerd op onderzoek door Schoenmaker (2012).

Waarop heeft het prestatiesysteem betrekking?

Het prestatiesysteem heeft betrekking op de kwaliteit en efficiëntie van het onderhoud van civiele infrastructuur. Dat wordt door beheerders als Rijkswaterstaat (RWS) in Nederland en de Britse Highway Agency (HA) uitbesteed aan marktpartijen. Daartoe worden prestatiecontracten opgesteld, waarin staat aangegeven welke prestaties de marktpartij moet leveren en welke vergoeding daar tegenover staat.

Soms is onderhoud heel eenvoudig (het maaien van bermgras bijvoorbeeld). Maar vaak heeft onderhoud een paar lastige kenmerken:

- Tegenstrijdige waarden: onderhoud moet bijvoorbeeld zo goedkoop mogelijk zijn, maar ook zo min mogelijk hinder voor de gebruiker veroorzaken. Tussen die twee waarden zit altijd een spanning.
- Afhankelijkheid van derden: de prestatie van de marktpartij kan ook sterk afhankelijk zijn van derde partijen – bijvoorbeeld andere overheden die toestemming moeten geven voor de onderhoudswerkzaamheden.
- Dynamiek: in de loop der jaren kunnen eisen vanuit de gebruiker of politiek, of technologische kennis veranderen, waardoor de prestatie-eisen verouderen.

Door deze variabelen is de relatie tussen output en outcome soms problematisch. Een marktpartij kan goede output leveren (door conform contract onderhoud te plegen), maar wanneer geen gebruik wordt gemaakt van de recente technologie of geen rekening wordt gehouden met nieuwe ontwikkelingen – kan de outcome niet gewenst zijn.

---

[10] Zouridis 2009, p.114.

De aanbestedende partij weet dit en staat daarmee voor een dilemma: gaat ze alles gedetailleerd voorschrijven, waardoor ze goed kan controleren of er geleverd wordt, maar niet flexibel is? Of biedt ze ruimte voor flexibiliteit, waardoor er altijd dispuut kan ontstaan over de vraag of de geleverde prestatie de gewenste prestatie is.

Wat is het doel of wat zijn de doelen van het systeem?

Doelen van het systeem zijn: 1) goed en efficiënt onderhoud; 2) inzicht in de prestaties van een marktpartij. Wanneer de marktpartij de afgesproken prestatie levert, krijgt ze de vooraf afgesproken vergoeding.

Wat is de typering van de prestaties die gemeten worden?

RWS gebruikt alleen output-indicatoren die iets zeggen over de kwaliteit van de elementen van de infrastructuur, maar die verder niet gekoppeld zijn aan de doelen van RWS. Denk hierbij aan indicatoren als wegonderhoud, reinigen van asfalt, beschikbaarheid en het percentage niet-werkende straatverlichting. De HA gebruikt daarnaast ook input- en throughput-indicatoren. Deze hebben tot doel aan te sluiten op de doelen van de HA en zijn meer gericht op leren en verbeteren. Bijvoorbeeld door het opstellen van gedetailleerde voorgeschreven activiteiten aan de hand van maintenance handbooks (input). Voorbeelden van throughout-indicatoren zijn de aanwezigheid van een kwaliteitssysteem en -plan, de impact van eerder behaalde scores (past performance data) en periodiek bijgestelde plannen van contractors. De snelheid van reageren op incidenten blijkt bijvoorbeeld perverse effecten op te roepen. Deze indicatoren zijn vooral bedoeld om te informeren.

Wat is de impact van het systeem?

Zowel bij RWS als bij de HA is sprake van een high impact systeem. De inkomsten van de marktpartij hangen af van het wel of niet nakomen van de gemaakte afspraken. De kwaliteit en efficiëntie van het onderhoud hangt af van het nakomen van de afspraken door de marktpartij.

Wat zijn de kenmerken van de groep die via het systeem wordt aangestuurd?

Ook hier geldt dat onder de prestatieafspraken een enorme variëteit schuil gaat aan onderhoudsobjecten: wegen, bruggen, sluizen. Dat is overigens niet heel problematisch: in de technische tradities van RWS en de HA past het om gedetailleerde overzichten te hebben van de onderhoudsopgave – in zogenaamde 'prestatiebestekken'.

Welke effecten van het systeem zijn gevonden?

- Bij zowel RWS als de HA leveren in veel gevallen de marktpartijen de afgesproken prestaties.
- Tegelijk hebben beide organisaties ook te maken met een grote verscheidenheid aan perverse effecten.[11] Denk hierbij aan inconsistent of verhullend rapporteren of het aanbieden van onvoldoende data, de neiging tot vervulling van de eisen van het contract in plaats van optimaal onderhoud, gebrek aan innovatie in onderhoud. Of denk aan het feit dat tijdens inschrijvingen wel wordt geconcurreerd op snelheid waarmee op incidenten wordt gereageerd, maar snel ter plaatse zijn betekent niet dat het incident daarmee sneller hoeft te worden opgelost. De beide organisaties ontwikkelen ook strategieën om die perverse effecten tegen te gaan.
- Er zijn ook verschillen in effecten tussen RWS en de HA:
  - RWS kiest voor een 'harde' houding. Vooraf definieer je de vereiste prestaties (ex-ante volledigheid), de marktpartij wordt geacht te leveren en daar reken je de marktpartij op af. Een relatief simpel beheersmodel dat onderhoud in vele kleine contracten opdeelt.

---

[11] Zie Schoenmakers (2011), p. 347-349.

- De HA biedt meer ruimte aan de marktpartij. Zij hanteert een beperkt aantal resultaateisen en veel werkvoorschriften: “bijvoorbeeld het zesmaandelijks preventief vervangen van lampen in drukknoppen bij oversteekplaatsen. Een grote mate van detail dus.” (Schoenmaker, 2011, p. 327). Bovendien zijn opdrachtgever en opdrachtnemer gezamenlijk verantwoordelijk voor prestatie-eisen als KPIs. Na sluiting van de overeenkomst bestaat ruimte om gewenste prestaties bij te stellen (ex post flexibiliteit). Een complexere vorm van beheersing dus. HA kiest er voor om ook tijdens het leveren van de prestatie, in contact te staan met de marktpartij – zo kunnen dilemma’s worden gedeeld, de impact van nieuwe technologische ontwikkelingen worden besproken, et cetera. Het gaat dan dus ook niet alleen om afrekenen – maar ook om leren.
- De benadering van de HA waarin dialoog en gezamenlijk leren en ontwikkelen domineren blijkt effectiever in termen van geleverde prestaties en heeft inmiddels ook bij RWS ingang gevonden. Uit het onderzoek van Schoenmaker wordt niet duidelijk wat de oorzaak is van die toegenomen effectiviteit.

Welke lessen kunnen uit de sector worden getrokken?

- Er is sprake van een grote hoeveelheid standaarden en voorschriften, die in combinatie met prestatie-eisen worden toegepast. Deze voorschriften bieden zekerheid en reduceren risico’s maar beperken de ruimte en vrijheid van de opdrachtnemer.
- De complexiteit en dynamiek van onderhoud impliceren dat onderhoud moeilijk is te vangen in een set indicatoren. Een desondanks precies gedefinieerde prestatie-indicator roept perverse effecten op.
- Hoe ‘harder’ de opstelling van de opdrachtgever (ex ante volledigheid, afstand, afrekenen), hoe meer en hoe groter de perverse effecten.
- Een houding waarin meer aandacht is voor en betrokkenheid bij het proces (flexibiliteit, interactie, ruimte voor leren) blijkt effectiever.

3.7. Off-shore gas

Beschrijving en analyse gebaseerd op onderzoek door Duiverman (2014).

Waarop heeft het prestatiesysteem betrekking?

Om het rendement van off-shore gaswinning te vergroten en de levensduur van boorplatforms te verlengen, worden boorplatforms op zee gemodificeerd. Hiervoor worden zogenaamde ‘mod’-projecten uitgevoerd. Het betreft technisch en logistiek complexe operaties die veel planning en voorbereiding vergen en waarbij vele disciplines (denk aan verschillende technische specialismen) betrokken zijn. Factoren als het weer op zee, beperkte ruimte om werkzaamheden uit te voeren op kleine constructies, hoge transportkosten en stringente veiligheidseisen spelen daarbij een grote rol. Bovendien worden modificaties gekenmerkt door een relatief hoge dynamiek. Dat wil zeggen dat gedurende een modificatietraject door de opdrachtgever nog vele veranderingen in de modificatie worden aangebracht (gemiddeld ongeveer 7 per project). Dit alles leidt tot projecten met een relatief hoge overhead in de vorm van projectmanagement en potentieel grote kostenoverschrijdingen en vertragingen in de daadwerkelijke uitvoering van modificaties. De kleine aantallen modificaties en hun uniciteit maakt een directe vergelijking van prestaties van modificatie-projecten lastig. Helemaal aangezien de modificatie van boorplatforms door gaswinningsbedrijven wordt uitbesteed. In de praktijk is zelfs sprake van een gehele keten van aannemers en onderaannemers. De complexiteit en wijze van organiseren creëert een principal-agent probleem. Het ontbreekt de opdrachtgever aan zicht op de prestaties. Opdrachtnemers beschikken niet over een systeem dat aantoont hoe goed zij presteren. In hoeverre zijn afwijkingen van inschattingen van projectkosten nu te wijten aan tussentijdse wijzigingen en hoe efficiënt zijn de aannemers? Er is de facto sprake van een efficiency- en verantwoordingvraagstuk.

Wat is het doel of wat zijn de doelen van het systeem?

Het prestatiemeetsysteem wordt gebruikt om de prestaties van modificatieprojecten in kaart te brengen. Deze informatie wordt gebruikt in de verantwoordingsgesprekken en contractonderhandelingen tussen gaswinningsbedrijven en (consortia van) aannemers.

Wat is de typering van de prestaties die gemeten worden?

Met behulp van een indicatorset die is gebaseerd op bestaande projectinformatie, dat wil zeggen informatie die aanwezig is voor projectregistratiedoeleinden, worden de prestaties van afgeronde projecten op een aantal essentiële criteria verzameld. De criteria hebben betrekking op de kernaspecten die in projectmanagement gemeten worden: 1) afgesproken kwaliteit, 2) planning en 3) budget (Atkinson, 1999). Omdat de factor kwaliteit een inherent subjectief criterium vormt, wordt kwaliteit als factor gesplitst in kwaliteit (ervaring klant) en de factor scope (voldoen aan *deliverables* of halen van projectdeadlines). Binnen de olie- en gasindustrie wordt gebruik gemaakt van een Project Management Triangle waarin wordt gefocust op de factoren tijd, kwaliteit en scope. De resultaten op deze dimensies bepalen de kosten van een project (Sylvester et al., 2008). Er wordt gebruik gemaakt van een kleine set prestatie-indicatoren (in totaal 13 prestatie-indicatoren). Op deze wijze kunnen de prestaties van (een consortium van) aannemers en onderaannemers in kaart worden gebracht.

Wat is de impact van het systeem?

De impact van het systeem is potentieel erg hoog. Ontevredenheid van de opdrachtgever (de gaswinningsorganisatie) over de prestaties van aannemers kan grote financiële consequenties hebben. Goede prestaties kunnen leiden tot verlenging van het onderhoudscontract; slechte prestaties tot beëindiging van het contract.

Wat zijn de kenmerken van de groep die via het systeem wordt aangestuurd?

Het prestatiemeetsysteem probeert inzicht te verkrijgen in een relatief kleine set afgeronde projecten. Uit de database van afgeronde projecten blijkt dat sprake is van een relatief kleine groep projecten, maar dat sprake is van een behoorlijke spreiding en variëteit in de wijze waarop modificaties worden uitgevoerd.

Welke effecten van het systeem zijn gevonden?

Het model is gebruikt in prestatieverantwoordingsgesprekken met gaswinningsbedrijven. Op basis van het model konden zorgen van het gaswinningsbedrijf worden geadresseerd.

De ontwikkeling van een laagdrempelig systeem dat gebruik maakt van managementinformatie uit de verschillende bedrijfsprocessen biedt managers de mogelijkheid om op gefundeerde wijze op de eigen prestaties te reflecteren. Als gevolg hiervan konden nieuwe inzichten worden gedeeld met de opdrachtgever.

Op basis van prestatiemeting kon worden vastgesteld dat in de 15 afgeronde projecten sprake was van een groot aantal (scope) wijzigingen, die grotendeels verantwoordelijk waren voor kostenoverschrijdingen en verhoogde managementkosten (gemiddeld 22%). Bovendien leidden de veranderingen tot een kostentoename van 207%. Op basis van het prestatiemeetsysteem kan worden geconcludeerd dat uitstel en een door de opdrachtgever geaccordeerde scope change ongeveer 45% van het verschil in uitkomsten kunnen verklaren.

Welke lessen kunnen uit de sector worden getrokken?

- Op basis van bestaande managementinformatie en/of data kunnen in een duidelijk afgebakende omgeving commerciële prestaties wel degelijk toerekenbaar worden aan specifieke ‘factoren’. Hierin lijkt sprake van een fundamenteel verschil ten opzichte van publieke prestaties. Het model laat zien dat door de klant gewenste veranderingen voor een belangrijk deel kostenoverschrijdingen kunnen verklaren. De duiding van deze informatie blijft lastig.

- *Ongoing* processen kunnen met behulp van prestatiemeting niet goed worden gemonitord en/of geëvalueerd. Het prestatiemeetsysteem kent daarmee een behoorlijke 'time-lag'. Lopende projecten kunnen niet worden meegenomen in de analyses.
- Voor zinvolle prestatiemeting is veel data vereist. In de database van contractors die voor een gaswinningsbedrijf alle activiteiten uitvoeren, is sprake van een relatief beperkt aantal projecten waardoor de relatieve meerwaarde van veel kwantitatieve indicatoren zeer beperkt is.

3.8. Veiligheidsdomein

In deze paragraaf beschrijven we recente ervaringen met prestatiemeetsystemen binnen de veiligheidswereld. Het is geenszins zo dat prestatie-indicatoren een nieuw onderwerp vormen in de veiligheidsregio's. In de loop der tijd zijn op het gebied van de veiligheidsregio's meerdere toezicht-, prestatie- en kwaliteitssystemen en -initiatieven ontwikkeld, die met behulp van indicatoren aspecten van presterend vermogen in kaart brengen.

Om de systemen en initiatieven in perspectief te plaatsen, beginnen we met een korte schets van de prestatievereisten voor veiligheidsregio's en het huidige toezicht daarop.

Toezicht door Inspectie VenJ

Sinds de inwerkingtreding van de Wvr in 2010 is de minister van Veiligheid en Justitie (voorheen minister BZK) verantwoordelijk voor het stelsel van brandweerzorg, geneeskundige hulpverlening in de regio, rampenbestrijding en crisisbeheersing (Algemene Rekenkamer, 2014). Om tot een goede uitvoering van de brandweerzorg en rampenbestrijding en crisisbeheersing te komen, definieert de wet normen die 'een uniform minimumniveau van veiligheid en van basiszorg' moeten bieden zonder dit begrip nader te specificeren. Wel staan in de wet en het Besluit veiligheidsregio's (Bvr) minimumeisen geformuleerd over de organisatie en de rol van de veiligheidsregio en de inrichting van de rampenbestrijding en meer specifieke prestatie-eisen over oefenen, alarmering, opschaling en opkomsttijden (input-eisen) (Van Veldhuisen et al., 2013; Algemene Rekenkamer, 2014; Inspectie VenJ, 2013, p. 20). Gesteld zou dan ook kunnen worden dat "prestatie-eisen (in termen van te bereiken resultaten) aan de rampenbestrijding en crisisbeheersing nagenoeg ontbreken" (Van Veldhuisen et al., 2013, p. 8).[12]

Via de Inspectie Veiligheid en Justitie ziet de minister dus sinds de vorming van de veiligheidsregio's toe op de wijze waarop veiligheidsregio's zich voorbereiden op rampen en crises en of wordt voldaan aan de wettelijke vereisten. De Inspectie beoordeelt de veiligheidsregio's met het wettelijk kader als uitgangspunt. Er wordt een beeld gegeven van het operationeel presterend vermogen van veiligheidsregio's. Op basis van de Wvr heeft de minister van VenJ de bevoegdheid om via de Commissaris van de Koning aanwijzingen te geven aan het bestuur van de veiligheidsregio als zij haar taken verwaarloost, maar dit wordt als een (te) zwaar instrument beschouwd en om die reden weinig ingezet (Van Veldhuisen et al., 2013, p. 11). De Inspectie maakt voor haar analyses gebruik van data die door de veiligheidsregio's uit (zelf)evaluaties worden aangeleverd, eigen inspectiegegevens en een ondersteunende documentstudie. Door informatie over voorbereiding op rampen en crises te spiegelen aan prestaties tijdens oefeningen en grootschalige incidenten (GRIP 3 of 4) komt de Inspectie tot een landelijke vergelijking van de prestaties van de veiligheidsregio's. Via de uitgave 'De Staat van

[12] De financiering van de veiligheidsregio is een gedeelde verantwoordelijkheid van de Rijksoverheid (via de Brede Doeluitkering) en het lokaal bestuur (gemeenten) In totaal gaat het om €1.2 mrd op jaarbasis, waarvan grofweg 15% wordt gefinancierd door het Rijk. (zie verder Algemene Rekenkamer 2014). Veiligheidsregio's worden dus overwegend lokaal en voor een klein deel landelijk gefinancierd. "De verhouding tussen lokale en landelijke financiering is echter niet in balans met de mate waarin de veiligheidsregio's worden gestuurd door lokale dan wel landelijke verplichtingen: in de praktijk worden taken, beleid en uitvoering van veiligheidsregio's grotendeels geregeerd door landelijk beleid (Van Veldhuisen et al. 2013, p. 9)."

de Rampenbestrijding' wordt door de Inspectie verslag gedaan van haar bevindingen, waarbij – zoals gezegd – de focus primair ligt op het al dan niet voldoen aan de wettelijke normen (Inspectie VenJ, 2013, p. 15-16). In haar inspectierapport identificeert de Inspectie koplopers en regio's die relatief achterblijven.[13]

Uit de verschenen rapportages en met name de laatste Staat van de Rampenbestrijding 2013 komt een dubbel beeld naar voren. Enerzijds wordt door de Inspectie, met name in de voorbereiding en inrichting van wettelijk vereiste hoofdstructuren, gewezen op een positieve trend in de voorbereiding op rampen en crises. Deze trend klinkt door in alle versies van de Staat van de Rampenbestrijding. Sinds de invoering van de Wet veiligheidsregio's en daaraan gekoppelde prestaties oordeelt de Inspectie dat "in het algemeen een vooruitgang zichtbaar is in de kwaliteit van de voorbereiding op rampen en crises. Anderzijds constateert de Inspectie dat geen van de veiligheidsregio's voldoet aan alle minimumeisen uit de wet (Van Veldhuisen et al., 2013, p. 9). "De veiligheidsregio's voldoen in toenemende mate aan de eisen die de Wet veiligheidsregio's op het gebied van rampenbestrijding en crisisbeheersing aan hen stelt", maar "de groei in het voldoen aan prestatie-eisen" als alarmering, opschaling en informatiemanagement is sinds 2010 "zeer beperkt" (Inspectie VenJ, 2013, p. 31). Ook voldoet geen van de veiligheidsregio's "aan alle voorwaarden die het Bvr hen oplegt" (Inspectie VenJ, 2013, p. 17).

Op basis van haar bevindingen rapporteert de Inspectie (2013, p. 22): "De veiligheidsregio's hebben tijdens de looptijd van dit onderzoek meer dan eens onder de aandacht gebracht dat de huidige wetgeving te veel is geformuleerd in termen van input. Tevens vinden zij de regelgeving overwegend te vormgericht, te technocratisch en te detaillistisch. Dit leidt ertoe dat nogal eens beargumenteerd wordt afgeweken van sommige bepalingen uit het Bvr." Kortom er wordt door zowel Inspectie als de veiligheidsregio's spanning ervaren tussen algemene wettelijke voorschriften en contextuele factoren en de roep om lokaal maatwerk. Er bestaat dan ook een "grote behoefte" aan een "nieuwe, concrete en op doelbereiking gebaseerde set van normen" (Inspectie VenJ, 2013, p. 29) en er wordt hardop gedacht over de vraag "of het naleven van de huidige wettelijke vereisten voldoende basis biedt voor een adequate rampen- en crisisorganisatie" (Van Veldhuisen et al., 2013, p. 9).

Aristoteles

In 2008 is getracht via het project Aristoteles te komen tot een stelsel van prestatie-indicatoren en een meetsysteem op outputniveau. Dit initiatief werd vanuit de brandweer en GHOR genomen en door de veiligheidsregio's via het Veiligheidsberaad overgenomen en had tot doel om een universele set van prestatie-indicatoren te formuleren op basis waarvan de prestaties van zowel de regio als de hulpdiensten (kolommen) *binnen* de veiligheidsregio's konden worden vastgesteld. Aristoteles had een aantal doelen: het creëren van transparantie in de prestaties van de veiligheidsregio, het kunnen vergelijken binnen de eigen regio en het bevorderen van de mogelijkheden voor leren. Daarnaast was een belangrijke drijfveer achter Aristoteles om de interne bedrijfsprocessen zoals de planning- en controlcyclus bij de veiligheidsregio's te introduceren en om managementinformatievoorziening te ontwikkelen en waar mogelijk te uniformeren. Primaire doelgroep van Aristoteles vormde de directie van de veiligheidsregio (Berenschot, 2009, p. 8).

Aristoteles heeft geleid tot zeven sets met een beperkt aantal prestatie-indicatoren voor de verschillende betrokken domeinen en kolommen (Brandweer, GHOR, Meldkamer, Crisisbeheersing & Rampenbestrijding, Bedrijfsvoering en directie en bestuursniveau). De indicatoren zouden bij voorkeur ingebed moeten worden in het INK managementmodel en kwaliteitssystemen van veiligheidsregio's. Tijdens het project Aristoteles werd geconcludeerd dat de organisatievormen van de veiligheidsregio's divers en complex zijn, waardoor geen

[13] Als gevolg van majeure stelselwijzigingen (invoering nieuwe wet Veiligheidsregio's (Wvr) en het besluit veiligheidsregio's (Bvr)) kan helaas geen langjarige trend worden onderkend. De Inspectie zelf spreekt van een "nulmeting" (IVenJ, 2013:17).

eenduidig landelijk beeld gemaakt kan worden van wat de prestaties van de veiligheidsregio's zouden moeten zijn. Daarnaast bleek tijdens het project dat landelijke wettelijke normen niet zo 'hard' zijn en dat veiligheidsregio's "op veel punten [...] regiospecifieke normen [...] stellen die afwijken van landelijke normen." (Berenschot, 2009, p. 5). Aristoteles ging uit van het idee van de ontwikkeling van een generieke set van indicatoren per veiligheidsregio, die aangevuld kon worden met specifieke indicatoren.

Tijdens de uitrol van Aristoteles, een verantwoordelijkheid van de veiligheidsregio's, bleek implementatie van de ontwikkelde indicatorensets in de praktijk lastig. Een deel van de regio's heeft de systematiek in meer of mindere mate overgenomen en sindsdien is het instrument verder ontwikkeld (Berenschot, 2014). Het project is in gewijzigde vorm doorgezet, waarbij niet langer alle prestaties werden ontwikkeld, maar de focus werd verlegd op de prestaties van de kolom brandweer. Dit onderdeel van het project Aristoteles werd in 2012 afgerond (Berenschot, 2014). Het herijkingsrapport formuleert de doelen van Aristoteles als de realisatie van "Een landelijk uniform model van prestatiemeting en -verantwoording" en de totstandkoming van "Een basisset van prestatie-indicatoren" (Berenschot, 2014, p. 9).

RemBrand

Op initiatief van het Veiligheidsberaad is onderzoek uitgevoerd naar criteria ten aanzien van brandveiligheid De opdracht aan de projectgroep luidde: "Benoem criteria (of indicatoren) die van invloed zijn op de brandveiligheid in Nederland. Werk dit uit in termen van output en outcome voor alle schakels van de veiligheidsketen en onderzoek hoe de schakels in de veiligheidsketen elkaar beïnvloeden en onderling kunnen worden gewogen." Het idee achter RemBrand was om meer balans te brengen in de benadering van brandveiligheid: het gaat niet alleen om de opkomsttijden en de voertuigbezetting. Tijdens de looptijd van het project zijn de beelden van de opdracht geëvolueerd: van een kwantitatief accent naar een kwalitatief accent en van een accent op besturing naar een accent op bevordering van brandveiligheid. De kwantitatieve benadering is verlaten omdat - mede vanuit ervaringen in het buitenland - blijkt dat de onderdelen van de veiligheidsketen zich niet zomaar kwantitatief laten verbinden. Er ontstaat daardoor een schijnzekerheid, mede door het idee dat met behulp van een "knoppenmodel" keuzes op een deel van de keten doorgerekend zouden kunnen worden op hun effecten op andere delen van de keten. Daarom is gekozen voor een primair kwalitatieve denklijn over brandveiligheid als geheel, waar de brandweer met haar opkomsttijden een fundamenteel onderdeel van uitmaakt. Een belangrijke drijfveer achter deze wijziging is het inzicht, dat brand een complex verschijnsel is, dat er geen standaardbrand bestaat en ook geen gesloten denkkader waarbinnen brand wordt voorkomen of kan worden beperkt (RemBrand, 2015, p. 7).

Cicero

Cicero heeft tot doel de kwaliteitszorg in de brandweer te verbeteren door middel van regionale 'peer reviews'– een gestructureerde vorm van intercollegiale toetsing (NVBR, 2010; Veiligheidsregio IJsselland, 2014). Cicero gaat uit van een vierjaarlijkse peer review. "Stip op de horizon is de visitatie van de veiligheidsregio, waarbij niet alleen de brandweer maar ook politie en GHOR (integraal) bevraagd worden op hun (multidisciplinaire) ambities en op prestaties in relatie tot vigerende normen." (NVBR, 2008, p. 10).

Het project Cicero bestaat uit vier verschillende deelprojecten, waarvan de laatste drie al afgerond zijn:

- Ontwikkeling van een visitatiesystematiek voor de veiligheidsregio
- Inrichting van een landelijk Kwaliteitsbureau Brandweer
- Ontwikkeling van een nieuw referentiekader voor de brandweer
- Ontwikkeling van een nieuwe auditsystematiek voor de brandweer

Visitatie

De Wvr (art. 56 lid1b) verplicht regio's om eens per 5 jaar een visitatie van de regio door een commissie te laten verrichten. Op dit moment wordt – als onderdeel van Cicero – gewerkt aan de ontwikkeling van een

visitatiesystematiek. In de tweede helft van 2014 is in de praktijk de methode van onderlinge visitaties beproefd in de regio's Kennemerland en Zeeland (zie tekstvak hierna). In 2015 en 2016 wordt het project vervolgd met een zestal regio's. "Met visitatie, in de vorm van peer review, wordt een strategische kwaliteitsfoto van de veiligheidsregio gemaakt: waar staat de organisatie op dit moment, welke doelen en ambities zijn al gerealiseerd en op welke punten kan de organisatie nog leren en verbeteren? De regio's houden zichzelf en elkaar als het ware een spiegel voor. Hierbij wordt vooral gekeken naar de verbeterdynamiek en de managementkracht van de organisatie (Ruijs, 2015, p. 13)."

Visitatiesystematiek pilotregio veiligheidsregio Kennemerland en Zeeland

De visitatiesystematiek is in het najaar van 2014 in een pilotsetting bij twee veiligheidsregio´s – Kennemerland en Zeeland – beproefd en verfijnd om te komen tot een definitieve opzet van visitatie die vanaf 2015 in alle regio's toegepast kan worden. In de visitaties zijn drie thema's behandeld: maatschappelijke opdracht van de veiligheidsregio, organisatie in samenwerking (in- en extern organiserend vermogen) en leiderschap en ontwikkelvermogen in de veiligheidsregio. De pilot had voor de twee pilotregio's als doel:

- Inzicht verkrijgen in punten waarop de samenwerking binnen en buiten de eigen veiligheidsregio kan verbeteren.
- Inzicht verkrijgen in de kwaliteit van de organisatie met betrekking tot de drie visitatiethema's.
- Inzicht verkrijgen in een realistische inschatting van de benodigde capaciteit, vooral intern, als het gaat om de voorbereiding van visitatie.

Commissie Hoekstra en Algemene Rekenkamer

Twee andere onderzoeken kunnen niet onvermeld blijven. Op verzoek van de minister van Veiligheid en Justitie heeft in 2012/2013 een onafhankelijke commissie onderzoek gedaan naar de werking van de Wvr en het brede Nederlandse stelsel van rampenbestrijding en crisisbeheersing. In september 2013 heeft deze Evaluatiecommissie Wet veiligheidsregio's en het stelsel van rampenbestrijding en crisisbeheersing (kortweg: de commissie Hoekstra) haar advies aan de regering uitgebracht. Ook vanuit de Algemene Rekenkamer (2014) is onderzoek gedaan naar de wijze waarop door veiligheidsregio's verantwoording wordt afgelegd over prestaties. Zowel de commissie Hoekstra als de Algemene Rekenkamer pleiten voor selectieve nationale sturing op een aantal onderdelen, waaronder de samenwerking tussen veiligheidsregio's onderling, tussen veiligheidsregio's en andere relevante organisaties, zoals de politie, en het hanteren van een kwaliteitszorgsysteem.

De Commissie Hoekstra baseerde haar advies mede op:

- De "Staat van de rampenbestrijding" van de Inspectie VenJ, met daarin de operationele prestaties van de veiligheidsregio's.
- De in opdracht van het WODC uitgevoerde evaluatie van de Wvr over het functioneren van het stelsel krachtens de Wvr en hoe actoren dat ervaren (Van Veldhuisen et al., 2013). Daarin wordt onder andere geconstateerd dat sprake is van vele processen in de wereld van de veiligheidsregio's die te maken hebben met presterend vermogen en sturing: "Parallel aan het verbeteren ten opzichte van de minimumeisen is in de afgelopen jaren gewerkt aan het formuleren van normen, eisen, standaarden en instrumenten; zowel binnen (de kolommen in de) veiligheidsregio's, als in landelijke structuren" (p. 13).

Over het toezicht inzake de rampenbestrijding en crisisbeheersing concludeerde de commissie Hoekstra dat het "effectiever" kan. De Inspectie VenJ, belast met het toezicht op de naleving van de Wvr, zou zich te veel richten op naleving van de wettelijke normen en te weinig op het feitelijk functioneren van de veiligheidsregio's. De Inspectie zou zich meer dienen te richten op kwalitatieve toetsing van plannen en risicoanalyses. Verder beveelt ze aan dat de Inspectie gaat werken met een een handhavingsstrategie die is gebaseerd op het principe van de toezichtsladder .

In de kabinetsreactie onderschrijft de minister van VenJ het advies van de commissie Hoekstra om het toezicht van de Inspectie VenJ meer te enten op het beoordelen van de operationele prestaties van de veiligheidsregio's en het opstellen van heldere normen daartoe. Ook het Veiligheidsberaad onderschrijft het advies van de commissie Hoekstra dat er nu te veel gestuurd wordt op organisatie-eisen en te weinig op effecten. Het Directeurenoverleg Veiligheidsregio's is van mening dat er meer sturing mag komen op het realiseren van gewenste maatschappelijke effecten in plaats van op inspanningen.

Volgens de Algemene Rekenkamer beschikt de minister over onvoldoende informatie om vast te kunnen stellen wat de veiligheidsregio's bijdragen aan het beschermen van burgers en bedrijven (Algemene Rekenkamer, 2014, p. 6). Volgens de Rekenkamer maakt het stelsel met 25 regio's het lastig om samenwerking en uitwisseling van kennis en informatie van de grond te krijgen en maken veiligheidsregio's nog onvoldoende gebruik van de mogelijkheden om kosten te besparen.

Kwaliteit en vergelijkbaarheid - Strategische Agenda Versterking Veiligheidsregio's 2014-2016

Naar aanleiding van de bevindingen van de commissie Hoekstra is door het Veiligheidsberaad een impuls gegeven aan het gezamenlijk vormgeven van kwaliteitszorg "om tot meer samenwerking, gezamenlijk leren en vergelijkbaarheid te komen" (Grijs, 2015, p. 13). Hiertoe is het project Kwaliteit en Vergelijkbaarheid opgesteld, dat tot doel heeft te komen tot "gezamenlijke vormgeving van kwaliteit en vergelijkbaarheid, het vergroten van het lerend vermogen van veiligheidsregio's en het versterken van de onderlinge samenwerking" (Grijs, 2015, p. 13). Met dit project wordt een nieuwe impuls gegeven aan initiatieven op het gebied van kwaliteit en vergelijkbaarheid en de bedoeling is om als schil om alle voorgaande initiatieven heen beschouwd te worden. Het doel wordt als volgt geformuleerd: "Kom tot het organiseren van integrale kwaliteitszorg, waarbij recht wordt gedaan aan de ontwikkelingen op het vlak van kwaliteit en vergelijkbaarheid in de regio's".

Incidentenonderzoeken

Een ongeval, ramp of crisis kan aanleiding zijn voor een incidentonderzoek. In sommige gevallen vindt dit plaats op verzoek van de minister van Veiligheid en Justitie. Soms zijn er meer onderzoeksinstanties betrokken bij incidentonderzoek, bijvoorbeeld de Onderzoeksraad voor Veiligheid en andere rijksinspecties. De aard van het incident bepaalt de opzet en invalshoek van incidentonderzoek. In de incidentenonderzoeken worden vragen beantwoord over onderwerpen als 'de verantwoordelijkheid van de veiligheidsregio ten aanzien van de voorbereiding van incidenten, de wijze waarop de regionale brandweer heeft geopereerd en de interregionale samenwerking'. De incidentenonderzoeken sluiten af met aanbevelingen aan de besturen van de veiligheidsregio om het presterend vermogen te vergroten.

Wij hebben drie van deze incidentenonderzoeken bestudeerd om factoren te identificeren die het presterend vermogen van veiligheidsregio's bepalen ('succesfactoren'). In overleg met de begeleidingscommissie zijn onderzoeken geselecteerd van verschillende typen incidenten van verschillende omvang die recentelijk (na invoering van de Wvr) plaatsvonden: een regio-overstijgend incident (GRIP[14] 4) en twee onderzoeken van incidenten van respectievelijk GRIP 2 en 3 formaat (die in de regel vaker voorkomen dan GRIP 4 incidenten). In alle onderzoeken zijn leerpunten geformuleerd voor de veiligheidsregio. De volgende incidentenonderzoeken zijn onderzocht:

[14] GRIP (Gecoördineerde Regionale Incidentenbestrijdingsprocedure) is een werkwijze die bepaalt hoe de samenwerking verloopt tussen alle hulpdiensten. De GRIP kent vier fases, afhankelijk van de grootte van het incidenten.

- Chemiepack Moerdijk (GRIP 4).[15]
- Aantreffen mosterdgas in een kelderbox in Ede (GRIP 3).[16]
- Brand in de Kelders in Leeuwarden (GRIP 2).[17]

In de incidentenonderzoeken zijn verschillende factoren geformuleerd die het presterend vermogen van een veiligheidsregio kunnen beïnvloeden:

Planvorming

- Vaststellen concept risicoprofielen en onderliggende documenten
- Aandacht in beleidsplan voor stille alarmering
- Opstellen gemeentelijk rampenplan
- Opstellen regionaal crisisplan
- Correct oefenbeleidsplan
- Oefeningen afgestemd op eventuele grote risico's
- Opstellen aanvullende regelingen in gebieden met grote risico's

Informatie en communicatie

- Informatie en communicatie ter plekke en in de omgeving
- Eenduidige afspraken over informatie-uitwisseling
- Afspraken over interregionale communicatie
- Eventueel tijdig openstellen van publieksinformatienummer en website
- Afweging open communicatie

Brandweerzorg

- Goed georganiseerde basisbrandweerzorg (rekening houdend met opkomsttijden en personele paraatheid)
- Eventuele aanwezigheid bedrijfsbrandweer
- Regie van de opkomst van brandweer
- Voldaan aan ondergrenswaarden brandveiligheid
- Tijdig duidelijkheid over aanwezigheid asbest
- Gebruik uitgangsstelling (een plaats waar hulpdiensten door de alarmcentrale naartoe worden gestuurd, voordat men een rampterrein op gaat)

Afstemming en verantwoordelijkheidsverdeling

- Heldere verantwoordelijkheidsverdeling
- Heldere bevelstructuur
- Heldere afstemming werkzaamheden
- Bekendheid met werkwijze partners

[15] Bij een brand in chemische fabriek op industriegebied Moerdijk in januari 2011 ontstond zeer zware rookontwikkeling richting de Drechtsteden. De Inspectie heeft het optreden van de crisisorganisatie van de veiligheidsregio onderzocht en meer in het bijzonder gekeken naar het optreden van de brandweer. Zie: http://www.ivenj.nl/actueel/inspectierapporten/rapport-brand-chemie-pack-moerdijk.aspx?cp=131&cs=64514

[16] Een hoeveelheid zeer gevaarlijke stoffen werd aangetroffen in een kelderbox van een appartementencomplex in de gemeente Ede in oktober 2013. Zie: http://www.ivenj.nl/actueel/inspectierapporten/rapport-evaluatie-grip3-incident-mosterdgas-ede.aspx?cp=131&cs=64514

[17] In de Kelders in Leeuwaren brak in oktober 2013 een felle brand uit. Zie: http://www.ivenj.nl/actueel/inspectierapporten/rapport-brand-aan-de-kelders-te-leeuwarden.aspx?cp=131&cs=64514

Opvang en verzorging

- Tijdige inrichting actiecentrum opvang en verzorging

GRIP-structuur

- Juiste toepassing GRIP-structuur
- Soepele overgang GRIP-fasen (bij op- en afschaling)
- Afspraken over taken in geval van afwijken GRIP-structuur

Wat opvalt is dat in de onderzoeken met name succesfactoren zijn geformuleerd die een procesmatig karakter hebben – ze hebben overwegend betrekking op de voorbereiding op een incident of procesafspraken rondom communicatie en verantwoordelijkheidsverdeling. Het kenmerk van dergelijke processen is dat ze zich veelal lastig laten uitdrukken in één of enkele indicatoren.

3.9. Inzichten uit andere sectoren

In dit hoofdstuk hebben we – op basis van een literatuurstudie – de ervaringen in verschillende sectoren met indicatoren voor het presterend vermogen op output- en/of outcomeniveau beschreven. In deze slotparagraaf gaan we in op de gemene deler. We identificeren een aantal generieke inzichten, die meegenomen worden bij het opstellen van ontwerpeisen voor een systeem van prestatiemeting voor de veiligheidsregio's (hoofdstuk 4):

- Prestatiemeting en de bijbehorende stelsels zijn complex. Er is sprake van een veelheid aan actoren en functies die aan prestatiemeting gekoppeld worden. In het veiligheidsdomein, maar ook in andere onderzochte sectoren vormt de toerekening van prestaties aan actoren een groot probleem. Zeker daar waar de outcome zeer moeilijk te preciseren is.
- In geen van de onderzochte sectoren zien we het gebruik van betekenisvolle outcome-indicatoren, er is sprake van hooguit output-indicatoren.
- Met uitzondering van het prestatiemeetsysteem in de gaswinning is geen enkele output-indicator of set van output-indicatoren in staat om de complexiteit van de geleverde prestatie te omvatten. De variëteit in de onderzochte sectoren laat zich niet gemakkelijk vangen in één of enkele indicatoren.
- Ervaringen met eerdere initiatieven op het gebied van prestatiemeting in het veiligheidsdomein (bijvoorbeeld Aristoteles en RemBrand) laten zien dat een zuiver kwantitatieve benadering van prestatie-indicatoren niet tot de gewenste resultaten leidt en 'schijnzekerheid' creëert. De implementatie van ontwikkelde indicatorensets is in de praktijk lastig gebleken. Recente initiatieven, bijvoorbeeld de ontwikkeling van een visitatiesystematiek, zijn overwegend kwalitatief van aard.
- Uit bestudering van incidentonderzoeken volgt dat factoren die bepalend zijn voor het presterend vermogen van de veiligheidsregio veelal een procesmatig karakter hebben en niet uit te drukken zijn in één of een beperkt aantal betekenisvolle indicatoren.
- Als de indicatoren in de prestatiemeetsystemen een hoge impact hebben, is het risico op perverse effecten eveneens hoog.
- In veel van de onderzochte sectoren zijn systemen 'doorontwikkeld'. Het lijkt er bijvoorbeeld op dat de aandacht verschuift van scores op indicatoren naar een gesprek met behulp van indicatoren.
- We hebben een aantal strategieën gezien om perverse effecten te beperken:
  - Besteed aandacht aan het proces van interactie tussen principaal en agent (gaswinning, zorg en onderhoud civiele infrastructuur). Geef het vertrouwen aan professionals dat degenen aan wie zij verantwoording afleggen de werkelijkheid 'on the ground' kennen (universiteit).
  - Beperk omvang en bereik van het systeem (zorg, politie en universiteit).
  - Beperk de impact van het systeem (alle casus).

De inzichten uit dit en het vorige hoofdstuk zullen wij in hoofdstuk 4 vertalen in een aantal ontwerpeisen, dat aan een systeem van prestatiesturing kan worden gesteld. Met die ontwerpeisen in de hand gaan we vervolgens naar de wereld van de veiligheidsregio's.

4. Ontwerpeisen

4.1. Inleiding

In dit hoofdstuk vatten we de bevindingen uit de voorgaande twee hoofdstukken samen in een aantal basale ontwerpeisen. De gedachte hierachter is dat een prestatiemeetsysteem ten minste aan deze eisen zal moeten voldoen, wil het betekenisvol en kansrijk zijn, en wil voorkomen worden dat de toepassing ervan leidt tot perverse effecten (bijvoorbeeld ‘afvinken’). We wijzen er met nadruk op dat dit basale ontwerpeisen zijn – als hieraan niet wordt voldaan, is het risico van pervertering van een prestatiemeetsysteem groot. Ze zijn een noodzakelijke, niet een voldoende voorwaarde.

We concretiseren de ontwerpeisen door per ontwerpeis een aantal vragen te formuleren die gebruikt kunnen worden om te achterhalen in welke mate aan de betreffende ontwerpeis wordt voldaan. In hoofdstuk 6 passen we de ontwerpeisen toe op de praktijk van de veiligheidsregio’s en komen we tot beantwoording van de centrale onderzoeksvraag.

4.2. Meetbaar

Ontwerpeis: De indicatoren in het systeem dienen zoveel mogelijk **meetbaar** te zijn.

Meetbaarheid heeft betrekking op de vraag of de prestatie eenduidig te meten is. En dus of aantoonbaar kan worden gemaakt dat sprake is van een verbetering of een verslechtering.

Om de meetbaarheid van een prestatie te bepalen, kunnen onder meer de volgende vragen worden gesteld:

- Is de prestatie waarneembaar? Of is de prestatie een *non-event*? Dat wil zeggen: iets dat <u>niet</u> is gebeurd.
- Is de prestatie kwantificeerbaar? Is de prestatie uit te drukken in een meetbare eenheid?
- Is de prestatie ondeelbaar? Of bestaat de prestatie uit meerdere deelprestaties?
- Als sprake is van deelprestaties, zijn deze dan onderling conflicterend? Dat wil zeggen: een goede score op de ene deelprestatie kan ten koste gaan van de score op de andere deelprestatie.
- Is sprake van een hoge mate van homogeniteit en uniformiteit in de sector waarop het prestatiemeetsysteem wordt toegepast? Of is sprake van variëteit?
- Hoe oordelen stakeholders over de meetbaarheid van de indicator?

4.3. Relevant

Ontwerpeis: De indicatoren dienen zo **relevant** mogelijk te zijn.

Bij relevantie draait het om de vraag of de prestatie uitdrukking geeft aan het bereiken van de doelstelling van de organisatie. Met andere woorden: doet de gemeten prestatie ertoe of niet?

Om de relevantie van een prestatie te bepalen, kunnen onder meer de volgende vragen worden gesteld:

- Betreft de prestatie een hoofdzaak? Of betreft het een bijzaak?
- Betekent een goede score op de indicator ook vrijwel zeker een goede prestatie? Of gaat achter de score op de indicator een complexe werkelijkheid schuil, waarin bijvoorbeeld professionele afwegingen een belangrijke rol spelen?
- Is de prestatie generiek en veelvoorkomend? Of bijzonder en uniek? Prestaties die bijzonder of uniek zijn, zeggen doorgaans minder over het presteren in het verleden en hebben ook een beperkte voorspellende waarde voor de toekomst.
- Hoe oordelen stakeholders over de relevantie van de indicator?

4.4. Toerekenbaar

Ontwerpeis: De indicatoren dienen zoveel mogelijk **toerekenbaar** te zijn.

Toerekenbaarheid heeft betrekking op de vraag of de prestatie die wordt gemeten toerekenbaar is aan degene die de prestatie op papier levert.

Om de toerekenbaarheid van een prestatie te bepalen, kunnen onder meer de volgende vragen worden gesteld:

- In welke mate is de prestatie het resultaat van de inspanning van de organisatie die wordt beoordeeld?
- Is de prestatie het resultaat van een samenwerking in een keten of netwerk?
- Welke organisaties zijn nog meer betrokken bij het leveren van de prestatie?
    - Hoe groot is de afhankelijkheid van deze andere organisaties?
- Welke niet-beïnvloedbare omgevingsfactoren zijn van invloed op het leveren van de prestatie?
    - Hoe groot is de invloed van deze niet-beïnvloedbare omgevingsfactoren?
- Hoe oordelen stakeholders over de toerekenbaarheid van de indicator?

4.5. Impact

Ontwerpeis: Naarmate meetbaarheid, relevantie en toerekenbaarheid meer problematisch zijn, dient de **gepercipieerde impact** van het prestatiemeetsysteem in de veiligheidsregio's beperkter te zijn om perverse effecten te voorkomen.

Of prestatiemeting leidt tot positieve effecten of tot perverse effecten hangt ook af van de impact van een systeem. Is die hoog of laag? Een hoge impact betekent dat het wel of niet behalen van een te meten prestatie grote gevolgen heeft in termen van verantwoording. Het gaat daarbij niet alleen om de feitelijke maar juist ook om de gepercipieerde impact: heeft het wel of niet behalen van een prestaties in de perceptie van de betrokken spelers een hoge impact?

Om de (gepercipieerde) impact van een prestatiemeetsysteem te bepalen, kunnen onder meer de volgende vragen worden gesteld:

- Hoe groot zijn de (potentiële) positieve effecten van het prestatiemeetsysteem volgens stakeholders, wanneer zij de afgesproken prestatie behalen?

- Hoe groot zijn de (potentiële) negatieve effecten van het prestatiemeetsysteem volgens stakeholders, wanneer zij de afgesproken prestatie niet behalen?
- Wat is functie van het prestatiemeetsysteem? Dient het om te leren of verantwoording af te leggen?
- Wordt het prestatiemeetsysteem gebruikt in een horizontale of verticale verantwoordingsrelatie?
- Bestaat er duidelijkheid over het prestatiemeetsysteem bij stakeholders? Of ervaren stakeholders onzekerheden?
    - Bestaat er duidelijkheid over de functie van het prestatiemeetsysteem bij stakeholders?
    - Bestaat er duidelijkheid over de wijze waarop prestaties worden geïnterpreteerd?
    - Bestaat er duidelijkheid over de wijze waarop de uitkomsten van het prestatiemeetsysteem worden gebruikt?
    - Is er duidelijkheid over de adressant(en) van het prestatiesysteem?
- Bestaat er draagvlak over de balans tussen kwantitatieve gegevens en kwalitatieve duiding in het prestatiemeetsysteem?
- Bestaat er draagvlak voor de wijze waarop het prestatiemeetsysteem wordt beheerd?

5. Beelden uit de regio's over prestatiemeting

5.1. Inleiding

De vraag of het mogelijk en wenselijk is om de prestaties van veiligheidsregio's meer te baseren op output- of zelfs outcome-indicatoren is niet slechts te beantwoorden met behulp van een analyse van literatuur over prestatiemeting of onderzoek naar andere instrumenten binnen en buiten de veiligheidsregio. Gegeven het feit dat het eventueel te ontwikkelen stelsel van indicatoren ingebed zal worden in een bestaande institutionele structuur, is de mate van draagvlak onder betrokken actoren van belang. De effectiviteit van een te ontwikkelen stelsel van indicatoren zal mede worden bepaald door de wijze waarop in de veiligheidsregio's draagvlak bestaat voor een dergelijke ontwikkeling. Om dit draagvlak te onderzoeken en te inventariseren welke ervaringen er reeds zijn met prestatie-indicatoren in bredere zin, hebben we gesprekken gevoerd met betrokkenen uit het domein van de veiligheidsregio.

Het draagvlak is in de volle breedte van het domein van de veiligheidsregio onderzocht, zowel op rijksniveau als op decentraal niveau. Allereerst zijn acht verkennende gesprekken gevoerd met zowel afgevaardigden van de Inspectie en de NCTV als directeuren van verschillende veiligheidsregio's. Vervolgens zijn op drie verschillende niveaus verdiepende gesprekken gevoerd in drie veiligheidsregio's, te weten: op bestuurlijk niveau (een burgemeester of voorzitter), op directieniveau en op operationeel niveau.[18] De beelden die uit deze gesprekken naar voren zijn gekomen, zijn getoetst in een bijeenkomst met afgevaardigden uit een aantal andere veiligheidsregio's. Dit leidde tot bevestiging van de beelden. Daarnaast hebben wij een systeemoefening bijgewoond in de regio Rotterdam-Rijnmond en ter plekke gesproken met betrokkenen uit de regio.

Uit deze gesprekken hebben we dominante beelden gedestilleerd die invloed hebben op de wijze waarop in de veiligheidsregio's nut en noodzaak van prestatiemeting van de veiligheidsregio worden ervaren. In dit hoofdstuk worden de beelden gepresenteerd. Dit hoofdstuk bevat alleen beelden die door meerdere gesprekspartners werden gedeeld, en geen beelden die een incidentele opmerking betroffen.

We geven hier geen ondertiteling bij of interpretatie van de beelden – die volgen in het volgende hoofdstuk.

5.2. Selectie van regio's

In samenspraak met de begeleidingscommissie zijn de volgende drie regio's geselecteerd voor het verdiepende onderzoek: Amsterdam-Amstelland, IJsselland, Gelderland-Zuid. Uitgangspunt is niet geweest om regio's te selecteren die een representatief beeld van de wereld van de veiligheidsregio geven. Een

[18] Voorbeelden van gesprekspartners op operationeel niveau zijn operationeel commandanten en stafmedewerkers met verantwoordelijkheden op het gebied van samenwerking of systeemoefeningen.

belangrijker criterium was dat de geselecteerde regio's reeds ervaring hebben met initiatieven die raakvlakken hebben met prestatiemeting. Zo zijn door één of meer van de drie regio's eigen evaluatiekaders opgesteld voor incidenten en wordt gewerkt met kennisdelingsgroepen, leerarena's, leeragentschappen en zelfevaluaties. Op basis daarvan is aangenomen dat gesprekspartners hun beelden baseren op reeds opgedane ervaring ten aanzien van prestatiemeting. Daarnaast is bij de selectie gekozen voor regio's die niet al veel eerder subject zijn geweest van pilots rondom visitaties of het onderzoek van de Algemene Rekenkamer. Bij de selectie van de regio's is tot slot rekening gehouden met een aantal overige criteria:

- De prestaties van de regio's: de regio's behoren alle drie tot de middenmoot, in de zin dat ze noch behoren tot het groepje dat in de Staat van de rampenbestrijding 2013 is aangemerkt als 'koplopers', noch tot het groepje dat is aangemerkt als 'achterblijvers'.
- De dichtheid van de regio's: de regio's verschillen van elkaar wat betreft het inwoneraantal en de bevolkingsdichtheid.
- Het risicoprofiel: de regio's verschillen van elkaar wat betreft risicoprofiel.

Hieronder volgt een korte beschrijving van de drie regio's met de specifieke kenmerken afkomstig uit de Staat van de rampenbestrijding 2013.

Profiel drie geselecteerde regio's

In de Staat van de rampenbestrijding 2013 behoren deze drie niet tot de koplopers of de achterblijvers, maar bij de grote groep middenmoters. Overigens is dit gebaseerd op een beperkt aantal items. Alle drie zijn zij bezig met een eigen traject om te komen tot prestatie-indicatoren.

Zowel IJsselland als Gelderland-Zuid zijn regio's met een mix van grote(re) steden en diverse kleine(re) gemeenten, van stedelijk gebied en platteland. Amsterdam-Amstelland is een randstadregio met een geheel eigen karakter. Amsterdam-Amstelland heeft op onderdelen een geheel eigen visie op hoe de veiligheidszorg in de praktijk gestalte zou moeten krijgen. Bovendien heeft Amsterdam-Amstelland veel ervaring met grootschalige evenementen en gebeurtenissen.

IJsselland

De Veiligheidsregio IJsselland heeft een half miljoen inwoners, dat is 3,1% van het totaal aantal inwoners in Nederland. De veiligheidsregio wordt gevormd door 11 gemeenten. De Veiligheidsregio IJsselland beslaat 4,6% van het Nederlands grondgebied en is daarmee relatief dunbevolkt. Een bovengemiddeld deel van het grondgebied kan worden getroffen door overstromingen. Er is veel vee in de regio, waarvan een bovengemiddeld aandeel runderen. Ook varkens, kippen en overig vee zijn in grote aantallen aanwezig. Eén van de achttien grote elektriciteitscentrales van Nederland ligt binnen de regio. Er liggen relatief veel strekkende meters wegen, spoorwegen en buisleidingen. Het aantal sluispassages voor de pleziervaart is relatief hoog, waarmee de (relatief weinige) waterwegen druk worden bevaren. Er zijn relatief weinig sluispassages voor de beroepsvaart. Een relatief klein deel van de regio kan worden getroffen door aardbevingen.

Amsterdam-Amstelland

De Veiligheidsregio Amsterdam-Amstelland is een regio met een grote variëteit aan kenmerken. De veiligheidsregio wordt gevormd door 6 gemeenten. Een hoge bevolkingsdichtheid gecombineerd met een aantal belangrijke toeristische trekpleisters die veel toeristen aantrekken, resulteren in de aanwezigheid van veel mensen in de regio. Zowel in Amsterdam als Amstelveen wonen veel verschillende nationaliteiten: in Amsterdam 178 nationaliteiten en in Amstelveen 124. Door effectieve overheidscommunicatie (bijvoorbeeld risico- en crisiscommunicatie) wil de regio borgen dat alle verschillende nationaliteiten bereikt worden. De regio ligt ingeklemd tussen zes veiligheidsregio's. Door de regio lopen diverse hoofdtransportroutes voor weg- en spoorvervoer, die zowel voor personen als goederen, waaronder ook

gevaarlijke stoffen, van groot belang zijn. Het Noordzeekanaal en het Amsterdam Rijnkanaal zijn belangrijke vaarwegen voor onder andere de zware industrie. De nabijheid van de luchthaven Schiphol betekent dat de regio ook rekening houdt met luchtvaartongevallen. In de regio worden ook grootschalige evenementen gehouden die veel publiek trekken (Sail, Koningsdag, Gay Pride, A Day at the Park, Ride for the Roses).

Gelderland-Zuid

De Veiligheidsregio Gelderland-Zuid beslaat een oppervlakte van circa 1.040 km². De ruim 530.000 inwoners wonen verspreid over achttien gemeenten. Deze gemeenten hebben op 8 juli 2010 gezamenlijk de Gemeenschappelijke Regeling Veiligheidsregio Gelderland-Zuid vastgesteld. Het gebied van de Veiligheidsregio Gelderland-Zuid kenmerkt zich door de aanwezigheid van drie belangrijke rivieren: Nederrijn, Waal en Maas, met de Waal als belangrijkste transportader over het water in Nederland. Het vervoer van gevaarlijke stoffen van Europoort naar Duitsland en vice versa vindt plaats over deze rivier. Daarnaast bevindt zich een aantal snelwegen (A15, A50, A73, A2) in de regio en worden er gevaarlijke stoffen vervoerd over het spoor via de Betuweroute. Ook kent de regio een aantal BRZO-bedrijven. In een aanzienlijk deel van de regio bestaat het risico van overstroming.

5.3. Onderzoeksprotocol

Op basis van de beelden uit de verkennende gesprekken is een protocol opgesteld voor de gesprekken in de verdiepende fase van het onderzoek. In het protocol is uitgewerkt op welk niveau van de veiligheidsregio gesprekspartners geselecteerd zijn en welke vragen hen zijn gesteld. Hierna wordt dit protocol weergegeven.

Protocol verdiepend onderzoek veiligheidsregio's

Doel onderzoek: Inzicht in de mogelijkheid en wenselijkheid om prestaties van veiligheidsregio's meer te baseren op output- en/of outcome-indicatoren.

Doel regiobezoeken: Het verzamelen van beelden op bestuurlijk niveau en in de praktijk met betrekking tot de vraag in hoeverre het mogelijk en wenselijk is om prestatie-indicatoren in veiligheidsregio's te gebruiken.

Uitgangspunten: Het onderzoek moet zich niet beperken tot bestaande prestatiemeetsystemen. We vragen daarom niet alleen naar ervaringen met bestaande indicatoren, maar ook – *green field* – naar gewenste prestaties: wanneer doet een regio het goed?

Wie?	*Gespreksonderwerpen*
Niveau 1: Bestuurders Burgemeester **Niveau 2: Directie** Directeur	**(Regio)specifieke context** • Vorming van de regio • Specifieke omstandigheden van de regio **Prestatiemeting** • Percepties over functioneren regio: o Wanneer bent u tevreden? o Wanneer zijn partners tevreden? o Wanneer zijn media en bevolking tevreden? • Prestaties regio: o Welke prestaties levert de veiligheidsregio? o Hoe verkrijgt u hier inzicht in? o Hoe stuurt u hierop? o Hoe weet u dat u in de koude fase voldoende bent voorbereid op de warme fase? • Prestaties kolommen: o Hoe vormt u zich hiervan een beeld? • Kwaliteit organisatie: o Hoe bepaalt u of de kwaliteit van processen/producten en materieel/mensen voldoende is? • Samenwerking: o Hoe verkrijgt u inzicht in samenwerking tussen kolommen en externe partners? • Inhoudsvolle verantwoording: o Hoe kunnen processen van verantwoording (richting gemeente, inspectie, etc.) zinvol worden ingericht?
Stap 3: Uitvoerders Medewerkers op operationeel niveau	**Prestatiemeting** • Percepties over functioneren regio: o Wanneer bent u tevreden? o Wanneer zijn partners tevreden? o Wanneer zijn media en bevolking tevreden? • Prestaties regio: o Welke prestaties levert de veiligheidsregio? o Hoe weet u dat u in de koude fase voldoende bent voorbereid op de warme fase? • Operationele processen: o Welke (operationele) processen zijn het meest bepalend voor de prestaties? • Kwaliteit organisatie: o Hoe weet u of de kwaliteit van processen/producten en materieel/mensen voldoende is?

Figuur 3 Onderzoeksprotocol

5.4. Beelden

Wij geven nu een overzicht van de dominante beelden uit de drie regiobezoeken. Eerst presenteren we een aantal inleidende beelden over het belang van prestaties en prestatie-indicatoren. Daarna werken we van min of meer contextuele beelden (over de bredere vraag naar de rol en verantwoordelijkheden van de regio's, bijvoorbeeld), naar de beelden die meer specifiek gaan over de betekenis van indicatoren. We hebben de beelden als volgt gerangschikt:

1. Over het belang van prestaties
2. Over de veiligheidsregio en verantwoordelijkheden
3. Over de zoektocht naar output en outcome
4. Over de relatie koud – warm[19]
5. Over het belang en de functies van prestatiemeting

5.4.1. Over het belang van prestatie

Het belang van inzicht in prestaties wordt erkend.
Alle gesprekspartners in de regio's onderkennen dat gegeven het maatschappelijk belang van veiligheid verwacht mag worden van veiligheidsregio's dat ze inzicht kunnen bieden in hun prestaties: de kwaliteit van organisatie en operatie, zowel in de koude als warme fase. Gesprekspartners zien in dat het belangrijk is om inzicht te krijgen in de toegevoegde waarde van veiligheidsregio's. En ze begrijpen de wens van de Inspectie VenJ en de NCTV om de effecten van veiligheidsregio's beter inzichtelijk te maken.

Het merendeel begrijpt ook dat de minister vanuit zijn rol als systeemverantwoordelijke inzicht dient te hebben in de kwaliteit van de taakuitvoering van de veiligheidsregio's en de zwakke plekken in het systeem. Gesprekspartners in de regio's vinden het bijvoorbeeld logisch dat de minister eist dat er een risicoprofiel ligt en vinden die eis goed passen bij de 'meta'-toezichtsrol die ze de minister toebedelen. Normen in de wet geven volgens gesprekspartners een beeld van *systeem*prestaties.

Aandacht voor randvoorwaarden en wettelijke vereisten is belangrijk.
Het huidige systeem van toezicht zoomt vooral in op randvoorwaarden en wettelijke vereisten uit de wet (zie eerst volgende tekstvak voor een beschrijving op hoofdlijnen van het toetsingskader van de Inspectie VenJ). Dat geldt met name voor de koude fase. Volgens gesprekspartners in de regio's beoordeelt het huidige toezicht 'dat de juiste boeken in de kast staan'. Gesprekspartners in de regio's onderkennen het belang van het op orde hebben van de randvoorwaarden. Bijvoorbeeld het hebben van een risicoprofiel, een crisisplan en het oefenen: "Tijdens die oefening kom je erachter dat 06-nummers zijn veranderd".

Maar kritiek op indicatoren: niet altijd de goede en weinig ruimte voor professionele afwegingen.
Er is echter ook kritiek op het toezicht door de Inspectie VenJ. Die kritiek betreft met name de prestaties waar het toezicht in de warme fase op controleert: die zijn voor veel gesprekspartners veel te gedetailleerd. Er ligt een te sterk accent op enkelvoudige, goed meetbare prestaties, zoals opkomsttijden. Die indicator zegt volgens gesprekspartners weinig over het feitelijk presteren van de betrokken functionarissen: "Soms kunnen hulpverleners beter eerst informatie inwinnen over de situatie ter plekke voordat ze gaan rijden, ook als ze daardoor de norm niet halen". De professional moet een afweging maken tussen twee waarden: op tijd ter plekke zijn of goed geïnformeerd ter plekke zijn en het systeem lijkt die afweging niet toe te staan.

[19] Koud betreft voorbereiding op veiligheidsincidenten en warm betreft het optreden tijdens een veiligheidsincident.

De Inspectie hanteert momenteel in haar toezicht een toetsingskader[20], waarin toetsingspunten zijn opgenomen. Dit toetsingskader biedt ondersteuning bij de evaluatie van incidenten en oefeningen. Het toetsingskader bestaat uit toetspunten die wettelijk verankerd zijn, ofwel in de Wet veiligheidsregio's ofwel in het Besluit veiligheidsregio's. Voorbeelden van wettelijk verankerde toetspunten zijn:

- Beschikt de veiligheidsregio over een actueel vastgesteld beleidsplan?
- Is er een actueel vastgesteld crisisplan?
- Fungeert de leider commando plaats incident als een discipline overstijgend (ontkleurd) leider?

Daarnaast zijn door de Inspectie in het toetsingskader nog aanvullende toetspunten opgenomen. De aanvullende toetspunten zijn niet op wettelijke vereisten gebaseerd, maar de Inspectie onderzoekt deze toetspunten wel en geeft hier in haar rapporten ook een oordeel over. Voorbeelden van aanvullende toetspunten zijn:

- Is in het risicoprofiel een overzicht opgenomen van risicovolle landsgrensoverschrijdende situaties?
- Is in het opleidings- en oefenbeleidsplan per (groep van) sleutelfunctionarissen de oefenfrequentie vastgelegd?

De laatste categorie toetspunten gaat in op onder andere de planvorming van de gemeenten, waardoor inzicht ontstaat in de (personele) gemeentelijke slagkracht. Voorbeelden van deze toetspunten zijn:

- Is de gemeentelijke rampenbestrijdingsorganisatie bij een incident in staat om voor de duur van de acute fase continu te kunnen functioneren?
- Is het team Opvang en Verzorging binnen 90 minuten begonnen met de werkzaamheden in de door de regio zelf vastgestelde minimale samenstelling?

Te veel status quo, te weinig ontwikkeling.

Gesprekspartners in de regio's geven aan dat in het toezicht nu een te sterke nadruk ligt op verantwoording en te weinig op leren. Deze nadruk wordt als ongebalanceerd ervaren. De Inspectie controleert volgens hen of er een crisisplan is, maar niet hoe het wordt toegepast, geëvalueerd en of ervan wordt geleerd: "De Inspectie meet nu status-quo en niet de ontwikkeling." In het verlengde hiervan: regio's willen meer flexibiliteit, bijvoorbeeld als het gaat om de vraag of het zinvol is om in dat jaar een systeemoefening uit te voeren.

Een ander voorbeeld van verantwoordingsdruk zijn de incident-onderzoeken van de Inspectie, die volgens gesprekspartners in de regio's geen goed beeld geven van wat er daadwerkelijk heeft plaatsgevonden en die voor de regio intern geen functie hebben.

De bovenstaande punten leiden tot kritische oordelen uit de regio over de prestatie-indicatoren in het huidige toezicht. De verantwoording richting de Inspectie wordt door gesprekspartners in de regio's in belangrijke mate ervaren als een "schijnwerkelijkheid", die inhoudelijk weinig bijdraagt aan de kwaliteit van de organisatie of het lerend vermogen van de veiligheidsregio's. Op basis van die 'schijnwerkelijkheid' worden regio's echter wel gerankt (bijvoorbeeld in de Staat van de rampenbestrijding). Die ranking heeft veel impact gehad. Sommige gesprekspartners geven zelfs aan dat die ranking 'buitenproportionele impact' heeft, omdat bestuurders er (veel) waarde aan toekennen ("bestuurders willen graag in het linkerrijtje"[21]). Vanwege de

---

[20] Inspectie Veiligheid en Justitie (2013) Toetsingskader staat van de rampenbestrijding. Den Haag: Inspectie Veiligheid en Justitie.
[21] Hiermee verwezen naar de beoordeling van de regio's in de 'Staat van de rampenbestrijding' en het belang dat bestuurders eraan hechten om goed beoordeeld te worden in vergelijking met andere regio's.

waarde die wordt toegekend aan de ranking, is het voor bestuurders van belang dat de beoordelingscriteria uit het toetsingskader - in ieder geval op papier - 'afgevinkt' worden.

5.4.2. Over de veiligheidsregio en verantwoordelijkheden

Regionale verantwoordelijkheid versus systeemverantwoordelijkheid Ministerie van VenJ: een worsteling.
Volgens gesprekspartners zijn veiligheid en voorbereiding op veiligheidsincidenten primair een regionale verantwoordelijkheid, die is belegd bij het regionale bevoegde gezag. Deze processen zijn regionaal en lokaal ingebed in planning en control cycli. Over deze processen wordt publieke verantwoording afgelegd aan het (verlengd) lokaal bestuur.

Tegelijkertijd erkennen gesprekspartners het belang en de betekenis van een nationale systeemverantwoordelijkheid. Veiligheidsregio's worstelen met de verhouding tussen die horizontale verantwoording (eigen regio) en de verticale verantwoording (minister). Als een regio door de Inspectie (verticaal) slecht beoordeeld wordt dan wordt de bestuurder in de regio ongerust, terwijl in de horizontale relatie vertrouwen was op basis van lokale verantwoordingsinstrumenten en -mechanismen. Het komt zelfs voor dat de bestuurder ongerust wordt als de regio te goed beoordeeld is, want dan bestaat het risico van bezuinigingen.

De Inspectie is zelf ook van mening dat de ranking een vertekend beeld kan geven van de praktijk. Ze benadrukt dat bij dit soort scores voorzichtigheid moet worden betracht.

Geen eenduidig beeld van de prestaties die veiligheidsregio's dienen te leveren.
Gesprekspartners geven geen eenduidig antwoord op de vraag welke prestaties veiligheidsregio's dienen te leveren – en welke indicatoren daarbij passen. Verschillende gesprekspartners formuleren wezenlijk andere antwoorden, die niet altijd alleen het gevolg zijn van de verschillende posities die gesprekspartners innemen. Sommige gesprekspartners benadrukken het belang van afstemming met andere regio's en nationale partners, andere benadrukken het belang van het faciliteren van leerprocessen, bijvoorbeeld het uitwisselen van kolomoverstijgende lessen met de kolommen en vice versa. Weer anderen geven aan dat de meerwaarde van een veiligheidsregio gelegen is in de aandacht op preventie.

De tabel hierna illustreert de uiteenlopende beelden onder gesprekspartners over de rol van veiligheidsregio's. Van drie gesprekken zijn de antwoorden weergegeven op de vraag wat de rol is van de veiligheidsregio. De verschillende type antwoorden zijn illustratief voor de verschillen in antwoorden in de overige gesprekken.

Gesprek A	Gesprek B	Gesprek C
• Hulp op straat • Rampenbestrijding en crisisbeheersing (voorbereiding en uitvoering) • Advies en toezicht • Platform- en regiefunctie	• Uitvoerende functie • Expertise- en adviesrol • Regie- en platformfunctie	• Gezamenlijke prestaties op initiatief van veiligheidsregio's • Interregionale samenwerking • Internationale samenwerking • Ketensamenwerking

Tabel 2. Beelden over rol veiligheidsregio.

Verschillende beelden over scheiding mono (kolommen) en multi (veiligheidsregio)
Sommige gesprekspartners zien sterke overlap tussen de prestaties van de veiligheidsregio en de prestaties van de kolommen. Om het presterend vermogen van veiligheidsregio's inzichtelijk te maken, moet volgens hen ook gekeken worden naar de prestaties van de kolommen. Als de prestaties in de kolommen goed zijn (inclusief de benodigde samenwerking) dan kan daaruit worden afgeleid dat ook de veiligheidsregio het goed doet.

Anderen geven aan dat de verantwoordelijkheid van de veiligheidsregio begint waar de verantwoordelijkheid van de kolommen eindigt. Zij benadrukken dat bij het inzichtelijk maken van het presterend vermogen van de veiligheidsregio's in de eerste plaats moet worden gekeken naar de kolomoverstijgende rol van de regio. Denk hierbij aan coördinatie tussen de kolommen en communicatie aan burgers over een ramp.

Met andere woorden: gesprekspartners in de regio's denken verschillend over de scheiding tussen mono (kolommen) en multi (veiligheidsregio). Dat betekent dat per regio andere prestaties aan de veiligheidsregio worden toegerekend. Als er sprake is van overlap tussen de prestaties van de veiligheidsregio en die van de kolommen, dan is het lastig om te bepalen welke prestaties dankzij de veiligheidsregio tot stand zijn gekomen.

Lastig om de toegevoegde waarde van de regio's te operationaliseren.
Wanneer tijdens de interviews gesproken werd over de prestaties van de veiligheidsregio kwam het gesprek doorgaans eerst op de prestaties van de kolommen en bleek het lastig om de toegevoegde waarde van de regio's te operationaliseren. Veel van wat een regio doet, bestaat uit afstemmen, samenwerken, elkaar informeren. Uitgerekend dit soort activiteiten is moeilijk te operationaliseren. Het gaat om een (1) relationele activiteit die (2) 'ongoing' is en (3) belangrijke, maar moeilijk te objectiveren effecten heeft. Bij prestatiesturing gaat het daarentegen om (1) producten, die (2) af te bakenen en (3) wel te objectiveren zijn.

Regio's zijn nog volop in ontwikkeling.
De Wet veiligheidsregio's is in 2010 in werking getreden. Veiligheidsregio's bestaan nog niet lang en zijn nog volop in ontwikkeling. Daardoor rijst bij enkele gesprekspartners de vraag of in deze fase de huidige toezichtwens van de Inspectie en de NCTV realistisch en wenselijk is. Daarbij is sprake van een duaal sentiment. Enerzijds leeft de perceptie in veiligheidsregio's dat veiligheidsregio's op zeer basale prestaties worden beoordeeld. Deze prestaties waren in het verleden wellicht goed om te definiëren om een ondergrens vast te stellen. In termen van afrekenen worden vragen gesteld bij de keuze voor deze normen. Anderzijds hebben regio's een ontwikkeling doorgemaakt, waardoor in termen van leren van de huidige normen weinig meerwaarde wordt ervaren. Gesprekspartners in de regio's zouden liever zien dat regio's de ruimte krijgen om aandacht te besteden aan het opzetten van een systeem van rampenbestrijding dat aansluit bij de specifieke context van de regio. De focus op een goede beoordeling door de Inspectie, leidt af van vragen over hoe het systeem van rampenbestrijding op lokaal niveau, gegeven een specifieke context, verbeterd kan worden.

Opvallend is overigens dat geen enkele regio ooit alle wettelijke vereisten gehaald heeft. Ook hier zien we strijdige beelden als we dit moeten interpreteren. Volgens enkele gesprekspartners roept het de vraag op of uitbreiding van het toezicht niet te vroeg komt – het toezicht zou moeten aansluiten op de noodzaak om in deze fase van de ontwikkeling vooral te leren. Anderen benadrukken dat het niet halen van wettelijke vereisten het belang bewijst van een prestatiemeetsysteem voor de wettelijke vereisten.

5.4.3. Over de zoektocht naar output en outcome

Tweeslachtig beeld.
Als de vraag naar output- en outcome-indicatoren wordt gesteld, zien we een tweeslachtig beeld. Enerzijds, output en outcome houdt regio's bezig – regio's denken in toenemende mate na over hoe output en outcome in beeld kunnen worden gebracht. Er lopen op dit vlak verschillende initiatieven. Bijvoorbeeld hoe beter inzicht kan worden verkregen in hoe burgers een ramp beleven. Anderzijds, gesprekspartners in de regio's wijzen er ook op dat eerdere initiatieven met prestatiemeting, zoals Aristoteles en RemBrand, op veel verzet stuitten. Ze betwijfelen of organisaties in dit stadium in staat zijn om (nog) nieuwe output- en outcome-indicatoren te absorberen.

Ingewikkelde worsteling rond thema uniformiteit – variëteit.
Rondom dit thema zien we een aantal beelden.

In de eerste plaats dat er 25 veiligheidsregio's zijn in Nederland. Tussen die regio's bestaat een zeer grote variëteit. Veiligheidsregio's verschillen in grootte, dichtheid, risicoprofiel, samenstelling van type gemeenten, ligging ten opzichte van grensstreken et cetera. Ook het type incidenten en veiligheidsvraagstukken waarvoor de veiligheidsregio zich gesteld ziet, is divers. Van stroomuitval tot onderzoek naar toekomstige veiligheidstrends, en bijvoorbeeld communicatie over risico's richting burgers. Bovendien wordt erop gewezen dat elk incident uniek is.

In de tweede plaats wordt er gewezen op de variëteit in de organisatie van veiligheidsregio's. Er bestaan naast overeenkomsten (gemeenschappelijke regeling, bestuur bestaande uit burgemeesters) ook verschillen tussen de regio's wat betreft de wijze waarop veiligheidsregio's zijn georganiseerd. Bijvoorbeeld: in meerdere regio's is een aparte organisatorische entiteit ingesteld met aan het hoofd een directeur; in andere regio's daarentegen is het veiligheidsbureau ondergebracht bij één van de grote gemeenten en zijn de taken van de directeur verdeeld over de hoofden van de kolommen. Veiligheidsregio's zijn een allesbehalve gelijkvormige bestuurslaag. Dat bemoeilijkt identificatie en totstandkoming van een eenduidig stelsel van prestatie-indicatoren. Tot slot verschilt de professionaliteit van de kolommen (vrijwilligers of professioneel) per regio.

Variëteit is dan ook veelvuldig genoemd als reden om voorzichtig te zijn met een set prestatie-indicatoren die groter is dan de set met wettelijke vereisten uit de regelgeving.

Tegelijk wordt ook het belang van uniformiteit genoemd. Partners, met name landelijke partners als ProRail en TenneT, hebben belang bij een zekere mate van uniformiteit. Landelijke partners hebben te maken met 25 regio's, die hun taak ieder op eigen wijze invullen. Bijna alle regio's onderscheiden gelijke risico's als pandemieën , ongevallen met gevaarlijke stoffen, hoogwater. De wijze waarop hieraan invulling wordt gegeven, verschilt echter in meer of mindere mate.

Integrale beeld wordt belemmerd doordat de politie niet meedoet.
Een probleem met toezicht op veiligheidsregio's dat ook organisatorisch van aard is, is dat voor alle regio's geldt dat de politie niet tot de veiligheidsregio behoort. Indicatoren voor de veiligheidsregio zouden de prestaties van de politie uitsluiten en dat belemmert volgens meerdere gesprekspartners een (min of meer) compleet beeld van prestaties van belangrijke partners in de rampenbestrijding en crisisbeheersing.

Wat belangrijke prestaties van veiligheidsregio's zijn, hangt af van wie je het vraagt.
In het veiligheidsdomein zijn verschillende type actoren actief: van bestuur, tot directie, tot operatie. Deze actoren hebben verschillende perspectieven en rollen en als gevolg daarvan uiteenlopende beelden van gewenste prestaties. Wat 'goede' prestaties zijn, is afhankelijk van aan wie je het vraagt. Een bestuurder vindt andere aspecten belangrijk dan de operatie: "operationeel gaat het om je best doen, bestuurlijk vooral hoe je scoort in rijtjes". In de tabel hierna worden door gesprekspartners genoemde voorbeelden van prestaties vanuit vier perspectieven beschreven. Deze rijtjes zijn niet limitatief.

Burger	Bestuurder	Directie / operatie	Partners
Tevredenheid over afhandeling ramp of crisis	Tevredenheid burgers	Tevredenheid burgers	Goede mensen
Minimaal aantal slachtoffers	Minimaal aantal slachtoffers	Kwaliteit organisatorische processen en procedures	Goed netwerk
Snelheid reactie	Snelle reactie	Goede mensen (vakbekwaam)	
Duidelijkheid / goede communicatie	Goede crisiscommunicatie	Goed netwerk ('weten bij wie je moet zijn')	
Goede nazorg	Snel, goed overzicht van situatie	Oefenen	
	Goede nazorg		
	Geen 'gedoe'		

Tabel 3. Verschillende beelden over type prestaties veiligheidsregio.

Burgertevredenheid wordt vaak genoemd als gewenste prestatie. De regio's zijn tenslotte in het leven geroepen voor de veiligheid van burgers. Bovendien: "De burger betaalt, dus mag hij weten wat daarvoor geleverd en verwacht mag worden". Gesprekspartners in de regio's geven aan dat burgertevredenheid onderbelicht is. Enkelen van hen geven aan dat ze het gevoel hebben dat regio's te veel aan het navelstaren zijn. Intern reflecteren wordt belangrijk gevonden, maar het is ook belangrijk om een beeld te hebben van de tevredenheid van de burger. Op het niveau van de kolommen wordt de beleving van burgers al onderzocht (bijvoorbeeld patiëntevaluatie en onderzoeken in de ambulance). Op het niveau van de veiligheidsregio wordt ook wel geëxperimenteerd met onderzoek naar burgertevredenheid, maar gesprekspartners geven aan dat de metingen nu nog vaak kwalitatief van aard zijn. Sommige gesprekspartners zijn van mening dat ten aanzien van burgertevredenheid wel 'harde' indicatoren te ontwerpen zijn, bijvoorbeeld: "wisten burgers tijdig waar ze 's avonds opgevangen zouden worden?".

Echter, gesprekspartners realiseren zich dat burgertevredenheid moeilijk te meten is. Gesprekspartners geven aan dat weinig burgers weten over wat de veiligheidsregio is en doet. Burgers herkennen waarschijnlijk vooral de prestaties van de kolommen. Tevens bestaat er een discrepantie tussen de beleving van de burger en de feitelijk prestaties van de regio of kolommen. Burgers weten niet wat er achter de schermen gaande is, bijvoorbeeld dat de brandweer ervoor kiest om een pand 'gecontroleerd' te laten uitbranden in plaats van het te blussen.

Regio's noemen output- en outcome-indicatoren voor de 'warme'-fase en input- en throughput-indicatoren voor de 'koude'-fase.

In de gesprekken is systematisch gevraagd naar prestaties van veiligheidsregio's en naar wat mogelijke output- en outcome-indicatoren kunnen zijn. Hierna volgt een groslijst van door gesprekspartners genoemde prestaties (uitgedrukt in indicatoren), ingedeeld naar type indicator (input, throughput, output, outcome) en naar fase ('koud' en 'warm'). Het betreffen voorbeelden; het overzicht is niet limitatief.

	Input	Throughput	Output	Outcome
Warm	• Aantal ingezette functionarissen • Kwaliteit ingezette functionarissen • Ingezet materieel • Kwaliteit ingezet materieel	• Kwaliteit bevelstructuur • Crisisleiderschap	• Reactietijd • Kwaliteit triage[22] • Duidelijkheid (kwantiteit en kwaliteit) crisiscommunicatie • Mate van ophef (voor bestuurders) • Kwaliteit nazorg • Juist geïnformeerde en gealarmeerde ziekenhuizen • Functioneren van verbindingen met partners, andere regio's en buitenland • Overgang gripfasen	• Aantal slachtoffers • Ernst van verwondingen • Omvang schade • Tevredenheid burgers over afhandeling crisis en dienstverlening
Koud	• Budget • Beschikbaar materieel • Beschikbaar aantal fte (per vakgebied) • Vakbekwaamheid mensen • Planning en controle systematiek • Capaciteit van andere regio's, ketenpartners en buurlanden • Regionaal crisisplan (organisatie, werkprocessen, taakverdeling) • Risicoprofiel	• Aantal crisisoefeningen • Kwaliteit crisisoefeningen • Afstemming oefeningen, plannen, regelingen op grote risico's • Netwerk met partners, andere regio's en buitenland ('weten bij wie je moet zijn') • Kennis van werkwijze partners • Aansluiting op meldkamer • Kwaliteit organisatorische processen en procedures • Uniformering werkprocessen door Veiligheidsberaad • Uniformering definities • Heldere verdeling taken (met name bestuurlijk) • Overgangsbevoegdheden gripfasen	• Voorspelbaarheid aanpak • Mate van leren • Duidelijkheid (kwantiteit en kwaliteit) risicocommunicatie	• Gevoel van veiligheid bij burgers

Tabel 4. In gesprekken genoemde voorbeelden van mogelijke indicatoren.

Ten aanzien van de genoemde prestaties kan een tweetal zaken worden opgemerkt. Ten eerste zijn er meer prestaties genoemd voor de 'koude' fase dan voor de 'warme' fase. Ten tweede zijn voor de 'warme' fase verhoudingsgewijs veel prestaties op output- en outcomeniveau genoemd, terwijl voor de 'koude' fase juist veelal prestaties op input- en throughputniveau zijn genoemd: "We weten of de poppetjes er staan, maar we weten niet wat ze presteren".

Output- en outcome-indicatoren zijn niet altijd goed meetbaar.

Sommige prestaties van veiligheidsregio's zijn goed meetbaar. Denk bijvoorbeeld aan het eerder genoemde voorbeeld van de opkomsttijden van functionarissen (nog even los van de constatering dat die indicator ook verhult). Voor andere prestaties geldt dat meetbaarheid problematisch is. Gesprekspartners wijzen erop dat de genoemde outcome-indicatoren vaak moeilijk objectiveerbaar zijn in één of enkele indicatoren (zeker als het woord 'kwaliteit' valt) en dat ze dus moeilijk meetbaar zijn. Wat is bijvoorbeeld 'kwaliteit van nazorg' als dat in een indicator moet worden neergelegd?

Bij output- en outome-indicatoren is er een sterke afhankelijkheid van externe factoren.

Gesprekspartners in de regio's wijzen er voorts op dat zich bij het leveren van output en outcome allerlei complicerende, externe factoren voordoen. Er zijn talloze variabelen van invloed op de output en outcome – en output- en outcome-cijfers zeggen nog weinig over de vraag of de regio het goed heeft gedaan. Zo is de impact van een incident niet alleen afhankelijk van het optreden van de veiligheidsregio, maar in het geval van bijvoorbeeld een ramp met het vrijkomen van gevaarlijke stoffen ook van de weersomstandigheden.

[22] Triage is het beoordelen van slachtoffers bij grote ongevallen, rampen en pandemieën, in verschillende categorieën verdeeld naar de ernst van de verwondingen of ziektebeeld.

5.4.4. Over de relatie koud – warm

De relatie koud – warm: beperkte betekenis output en outcome tijdens warm.
De relatie koud – warm is in de gesprekken veelvuldig besproken. Koud betreft voorbereiding op veiligheidsincidenten en warm betreft het optreden tijdens een veiligheidsincident.

De zeggingskracht van outcome- en output-indicatoren in de warme fase is beperkt vanwege de lage frequentie en het unieke karakter van een ramp/crisis. GRIP 1-optreden vinden ongeveer zes keer op jaarbasis plaats in een gemiddelde veiligheidsregio, en hoe hoger de GRIP-fase hoe zeldzamer het incident. Van een ramp/crisis wordt vaak een uitgebreide evaluatie gemaakt, die vele malen rijker is dan een eenvoudige set indicatoren.

De relatie koud - warm is niet eenduidig.
Een alternatief is prestaties meten in de koude fase. Echter de relatie koud - warm is niet eenduidig. Hoe voorbereiding op optreden en daadwerkelijk optreden zich tot elkaar verhouden, is niet altijd duidelijk. Koud is niet altijd een goede voorspeller voor warm. Dat betekent dat ook de prestaties in de koude fase beperkte zeggingskracht hebben voor de prestatie van de veiligheidsregio. Prestaties in de koude fase hebben bovendien – zoals ook blijkt uit de tabel hiervoor – eerder het karakter van een input- of procesindicator in plaats van output of outcome.

Tegenover deze kritiek staat dat de vragen die over de relatie koud-warm worden gesteld, als legitiem worden ervaren. Bijvoorbeeld de vraag 'hoe weet de regio in de koude fase dat ze goed is voorbereid op de warme fase', wordt als een zeer relevante vraag bestempeld, die beantwoord moet worden – de zorg is alleen dat deze vraag wordt gereduceerd tot een vraag naar indicatoren.

De zoektocht naar output- en outcome-indicatoren voor preventie komt ook langszij.
In de gesprekken is duidelijk geworden dat de aandacht van de veiligheidsregio's in toenemende mate verschuift naar de voorkant van de veiligheidsketen.[23] Gesprekspartners zijn het erover eens dat hier relatief gezien veel meer veiligheidswinst valt te behalen. Dus naast crisisbeheersing ook aandacht voor het voorkomen van incidenten (preventie) en de zelfredzaamheid van burgers, zodat zij beter voorbereid zijn als een incident plaatsvindt. De discussie over de meetbaarheid over dit soort prestaties staat nog aan het begin.

Gesprekspartners in de regio's hebben twee constateringen. Ten eerste, deze prestaties zijn minder goed meetbaar. Bij preventie speelt de vraag: hoe meet je iets dat niet heeft plaatsgevonden? Ten tweede, er is wel een aantal relatief eenvoudig te operationaliseren indicatoren – bijvoorbeeld in het geval van zelfredzaamheid: 'het percentage bewoners dat thuis een brandblusser heeft' – maar dan dringt zich de vraag op of dit niet een te vereenvoudigde weergave van de realiteit is. Elke poging om meervoudige prestaties als preventie en zelfredzaamheid uit te drukken in één of enkele indicatoren betekent dat andere aspecten (onterecht) buiten beschouwing worden gelaten. Het risico bestaat dat de aandacht uitgaat naar die activiteiten die wel goed meetbaar zijn.

Output en outcome-indicatoren: ingebed in verantwoordelijkheidsdiscussie.
Sommige gesprekspartners vragen zich af waar de behoefte aan output- en outcome-indicatoren vandaan komt. Zij vinden die wens niet passen bij de systeemverantwoordelijkheid van de Rijksoverheid. Volgens hen bestaan er in de relatie tussen gemeenteraad en veiligheidsregio al afdoende verantwoordingsinstrumenten en -mechanismen. Bovendien roept de huidige toezichtsvraag van de Inspectie en de NCTV de fundamentele

[23] De veiligheidsketen bestaat uit de volgende processen: proactie, preventie, preparatie, repressie, nazorg.

discussie op waar de verantwoordelijkheid voor veiligheid in brede zin ligt. Volgens gesprekspartners ligt deze discussie gevoelig, omdat deze aan de decentrale autonomie raakt. De beelden over de wenselijkheid van output- en outcome-indicatoren worden in sterke mate gekleurd door de discussie ten aanzien van systeemverantwoordelijkheid (van de minister van VenJ) en lokale verantwoordelijkheid rond veiligheid. Waar eindigt de een en begint de ander?

Normatieve karakter van indicatoren.

Gesprekspartners geven aan dat met prestatiemeting wordt geprobeerd om te objectiveren, maar dat dit juist rond veiligheid moeilijk is. Want welk aantal slachtoffers is acceptabel? En is het voldoende als je een 7 scoort op de beleving van veiligheid van bewoners?

Bij spoorveiligheid zie je dit bijvoorbeeld ook: daar zijn streefcijfers voor het aantal slachtoffers, maar die spelen in de (politieke) discussies geen rol want elke dode is er één te veel, zelfs al is er sprake van een dalende trend in het aantal slachtoffers.

Nieuwe indicatoren (alleen) zijn niet de oplossing.

Het huidige toezichtstelsel meet volgens betrokkenen niet de juiste prestaties en/of hanteert niet de juiste indicatoren om inzicht te krijgen in de prestaties. Nieuwe indicatoren worden echter veelal niet als de 'way out gezien' – het gaat om meer dan alleen dat. Hierbij wordt ook gewezen op eerdere pogingen om prestaties van veiligheidsregio's te kwantificeren. Zo is binnen het project Aristoteles een indicatorensysteem ontwikkeld. Dat systeem had tot doel om de prestaties van zowel de regio als de hulpdiensten (kolommen) *binnen* de veiligheidsregio's in beeld te brengen. Het systeem is volgens enkele gesprekspartners niet geslaagd: hun beeld is dat er maar weinig veiligheidsregio's over zijn gegaan tot implementatie van (delen van) het systeem.

Andere geluiden zijn in de gesprekken ook geuit. Nieuwe indicatoren die beter uitdrukking geven aan de prestaties van veiligheidsregio's worden niet op voorhand afgewezen. "Er zitten veel haantjes in deze organisatie. Competitiedrang is goed als instrument. Maar dan wil ik wel de juiste hulpmiddelen hebben." Tegelijk geldt dat veel van de voorbeelden van indicatoren die worden genoemd hun tekortkomingen hebben (zie de tabel 3 hierboven en de observaties naar aanleiding van deze tabel).

Het belang van een kwalitatief goed gesprek tussen het ministerie en de regio, waarin aan de hand van slimme vragen verdieping wordt aangebracht op de prestatie-indicatoren, wordt breed gedragen. Een dergelijk gesprek geeft meer inzicht in de werkelijkheid achter het cijfer – meer dan het toevoegen van een aantal extra indicatoren.

Angst voor te sterke sturende werking output en outcome in warme fase.

Veiligheid wordt onder normale omstandigheden (in de koude fase) als vanzelfsprekend bevonden – veiligheid is de norm. Wanneer zich echter een incident voordoet, gaat dit veelal gepaard met veel emoties: "Een GRIP 3 ramp gebeurt maar één keer per twee of drie jaar en dan staan alle camera's erop en ligt alles onder een vergrootglas". Media, Rijk en bestuurders zitten er heel dicht bovenop als het 'fout' gaat.

In de gesprekken is de angst uitgesproken dat output- en outcome-indicatoren, die zowel onthullen als verhullen, verkeerd geïnterpreteerd zullen worden. Men is bang voor disproportionele maatregelen op basis van de cijfers.

Sommige gesprekspartners wijzen hierbij ook op de risicoregelreflex: na een incident komen er vaak nieuwe regels bij. Media en burgers hebben volgens hen vooral oog voor als het fout gaat, en staan minder vaak stil bij geslaagde systeemoefeningen of incidenten die goed zijn afgehandeld (of zijn daarin minder geïnteresseerd).

5.4.5. Over het belang van en de functies van prestatiemeting

Bestuurders hechten veel belang aan prestatiemeting.

Eerder beschreven we al dat de uitkomsten van het huidige toezicht van groot belang worden bevonden door bestuurders. Zo heeft de ranking in de Staat van de rampenbestrijding van de Inspectie VenJ volgens gesprekspartners buitenproportionele impact, omdat bestuurders er (veel) waarde aan toekennen ("Er is geen bestuurder die de laatste wil zijn op de lijstjes van de Inspectie").

Als de nadruk in het prestatiemeetsysteem verschuift naar output en outcome zal de impact naar verwachting verder toenemen. De nadruk zal immers komen te liggen op die zaken die de aandacht zullen trekken van media, bewoners en politiek. De belangen voor bestuurders worden daardoor groter. "Een crisis kan een bestuurder maken of breken." En: "Veiligheid is voor bestuurders een afbreukdossier."

Leren versus afrekenen.

Een prestatiemeetsysteem is een instrument dat, zoals eerder is geobserveerd, bestuurlijk zwaar wordt gewaardeerd. Gesprekspartners in de regio's zijn bang dat het instrument gebruikt gaat worden om af te rekenen. Afrekenen op indicatoren die maar een deel van de echte werkelijkheid representeren, is risicovol en kan tot pervertering leiden. De voorbeelden hebben we hierboven genoemd: afvinken en bureaucratie, verhullende werking prestatie-indicatoren en 'gaming the numbers' (zie hoofdstuk 2).

Overigens is het soms moeilijk om aan te geven of een effect positief is of pervers. Het voorbeeld is genoemd van een regio waarvan het bestuur bewust ervoor koos om geen systeemoefening te doen. "Dat leidde tot een 'foei' van de Inspectie in het rapport [Staat van de rampenbestrijding]. Er gaat een afschrift naar de commissaris van de Koning. Achteraf is door het bestuur gezegd dat in het vervolg misschien toch maar gewoon geoefend moet worden." Dit kan als een positief effect van prestatiesturing worden geduid: er wordt gewoon weer een systeemoefening gehouden. Maar het kan ook als een pervers effecten worden geduid – de regio ziet de meerwaarde van een systeemoefening niet en wordt daartoe toch gedwongen door het systeem.

Veel aandacht voor leren en evalueren.

Veel regio's willen leren – en willen evaluaties gebruiken om te leren. Er zijn zelfs regio's die vinden dat presterend vermogen gedefinieerd kan worden als de capaciteit om leerprocessen goed te organiseren: "Wat in multi-verband geleerd wordt, moet in mono teruggekoppeld worden. En andersom. De veiligheidsregio moet zicht hebben op leerpunten in de kolommen." En: "Welke rol speelt de regio om in de kolom te prikkelen?". Om zicht te krijgen op prestatie van regio's, stellen regio's voor om de vraag te stellen: "Hoe zijn regio's omgegaan met de leerpunten uit de incidenten (oefeningen)?".

In de gesprekken is genoemd dat er op dit moment binnen de veiligheidsregio's al veel aandacht is voor leren en evalueren. Naast landelijke initiatieven, zoals het ontwikkelen van een gezamenlijke visitatiesystematiek en het project Kwaliteit en vergelijkbaarheid, zijn er diverse regionale en interregionale initiatieven. Zo zijn er kennisdelingsgroepen, bijvoorbeeld in Oost-Nederland met drie veiligheidsregio's uit Gelderland en twee uit Overijssel. Ander voorbeelden die zijn genoemd, zijn: leerarena's, leeragentschappen en zelfevaluaties.

Door meerdere gesprekspartners is ook aangegeven dat leren en evalueren nog verder doorontwikkeld kan worden binnen het veiligheidsdomein. Er wordt vooral geëvalueerd en geleerd naar aanleiding van incidenten (want die zijn landelijk het meest in beeld). Dat is volgens deze gesprekspartners niet voldoende. Leren zou integraal onderdeel moeten uitmaken van alle activiteiten van een regio. Bijvoorbeeld ook van preventie, eveneens een belangrijke taak van de regio.

Er is dus veel aandacht voor leren. Tegelijkertijd geldt: de activiteiten die daartoe worden ontplooid, laten zich moeilijk in betekenisvolle indicatoren vangen.

Veiligheidsregio's doen zelf actief aan kwaliteitsborging.
Zowel gezamenlijk als op individueel niveau denken regio's na over het introduceren en doorontwikkelen van systemen voor kwaliteitsborging. De wereld van de veiligheidsregio kent meerdere van dergelijke systemen en mechanismen. In de tabel hieronder staan de belangrijkste kort beschreven.

	Toezicht Inspectie VenJ	Gesprekken NCTV	Kwaliteitszorg(systemen)	Visitatie	Kostenevaluatie
Functie	Verantwoorden	Monitoren, kennis en informatiedeling	Leren en verantwoorden	Leren	Leren
Richting	Verticaal	Verticaal	Horizontaal	Horizontaal[24]	Horizontaal
Frequentie	Doorlopend	Jaarlijks	Continu	Vijfjaarlijks[25]	Driejaarlijks
Ontwikkelingen			Implementatie Aristoteles Project Kwaliteit en Vergelijkbaarheid[26]		

Tabel 5. Initiatieven van systemen voor kwaliteitsborging.

In de gesprekken is genoemd dat een stelsel van prestatie-indicatoren ten minste op de bestaande systemen zou moeten aansluiten.

Veel draagvlak voor visitatie.
Uit de gesprekken volgt het beeld dat er veel draagvlak is in de regio's voor visitaties. De verwachtingen bij de gesprekspartners zijn hoog: "Collegiale toetsing en visitaties, dat helpt mij oprecht", "Het is leerzaam als je bij collega's in de keuken gaat kijken", "Collegiaal visitatie van mensen die doorvragen, dat helpt".

Wel is opgemerkt dat de huidige visitatiesystematiek nog in de kinderschoenen staat en verder moet worden ontwikkeld om als volwaardig systeem voor kwaliteitsborging te kunnen fungeren.

5.5. Andere zienswijzen

Het beeld dat uit de voorgaande paragraaf naar voren komt is dat er veel twijfel is over de toegevoegde waarde van output- en outcome-indicatoren. Die twijfel is ingebed in beschouwingen over (1) de verschillen tussen veiligheidsregio's, (2) de (gepercipieerde) spanning tussen verticale en horizontale verantwoording en (3) de dynamiek van de regio's, die vaak nog eenheden-in-ontwikkeling zijn.

Op de vraag hoe we deze beelden moeten waarderen, komen we terug in het volgende hoofdstuk. Voor nu is belangrijk dat er ook andere zienswijzen zijn, die ons deels vanuit de regio, deels vanuit de Inspectie en de NCTV zijn aangereikt. Wij vatten deze kort samen:

[24] Het bestuur van de veiligheidsregio is wel verplicht om het rapport van de visitatiecommissie toe te zenden aan de minister.
[25] In artikel 56 van de Wvr wordt de verplichting opgelegd aan het bestuur van de veiligheidsregio om elke 5 jaar een visitatie uit te laten voeren door een visitatiecommissie.
[26] Doel van het project is: "het vergroten van het inzicht in de kwaliteit van de regio's, het verbeteren van het lerend vermogen, het bevorderen van multidisciplinaire samenwerking en het vergroten van de onderlinge samenwerking tussen regio's". Op basis van de doelstelling is de volgende opdracht gedefinieerd: "Kom tot het organiseren van integrale kwaliteitszorg, waarbij recht wordt gedaan aan de ontwikkelingen op het vlak van kwaliteit en vergelijkbaarheid in de regio's".

- De roep om betere indicatoren dan de huidige proces-indicatoren kwam ook vanuit de regio (Veiligheidsberaad). Zo ontstaat een wat diffuus beeld: enerzijds is er vanuit de regio de roep om betere indicatoren, anderzijds wordt dit desgevraagd niet concreet gemaakt en wordt door regio's vooral gewezen op de nadelen van indicatoren.
- Bij het beeld dat de veiligheidsregio's nieuw zijn en dus nog tijd nodig hebben om zich te ontwikkelen, passen twee kanttekeningen: de veiligheidsregio's bestaan al bijna 5 jaar (sinds 2010) en zijn niet uit het niets ontstaan. Het is dan ook de vraag hoe lang de regio's zich er op kunnen beroepen dat ze zich nog in een ontwikkelfase bevinden.
- Ten aanzien van de worsteling die regio's ervaren tussen horizontale en verticale verantwoording kan worden opgemerkt dat de onderzoeken van de Inspectie juist ook een bijdrage leveren aan de horizontale verantwoording. Binnen de wettelijke kaders bepalen de besturen van de verschillende veiligheidsregio hun eigen ambities en het kwaliteitsniveau van de veiligheidsregio. Dat betekent dat de veiligheidsregio's de vormgeving van hun taken verschillend kunnen invullen. Die verschillen mogen er echter niet toe leiden, dat de bovenregionale samenwerking bij incidenten in gevaar komt. Daarop toezien is een gedeelde verantwoordelijkheid van de gemeenten en het ministerie van VenJ.
- Bij het beeld dat er uiteenlopende beelden zijn over de prestaties van de veiligheidsregio kan de kanttekening worden geplaats dat in de wet diverse taken zijn vastgelegd. Naast deze wettelijke taken kunnen regio's aanvullende taken oppakken.
- Ook bij het beeld dat de Inspectie slechts de status-quo beschouwt en niet naar ontwikkelingen kijkt, zijn kanttekeningen te plaatsen. In de Staat van de Rampenbestrijding 2013 wordt aandacht besteed aan lopende (verbeter)trajecten en ontwikkelingen.
- Er is vanuit de regio's kritiek op de ranking. Wederom: ook hier is een concurrerend beeld mogelijk. Rankings hebben op z'n minst bestuurders geactiveerd en zo tot activiteit en beweging geleid. De drang om in het 'linkerrijtje' terecht te komen, kan ook positief worden geduid.
- Hoewel de politie niet tot de verantwoordelijkheid van het bestuur van de veiligheidsregio's behoort maakt zij wel afspraken (samenwerkingsconvenant) over de door de politie te leveren prestaties ten aanzien van de (voorbereiding op) een ramp of crisis.

Dergelijke beelden leiden tot een centrale observatie die we vaker tegenkomen: er zijn steeds twee beelden van eenzelfde werkelijkheid, die allebei iets legitiems hebben. De onderliggende verklaring wordt door vrijwel iedereen erkend: er is een spanning…

- …tussen horizontale en verticale verantwoordingsrelaties;
- …tussen afrekenen met indicatoren en leren van indicatoren;
- …tussen de complexe en meervoudige werkelijkheid en de eenduidige en eenvoudige indicator.

In het volgende hoofdstuk gaan we in op de vraag hoe we dit alles wegen om tot een uitspraak te komen over de wenselijkheid en mogelijkheid van output- en outcome-indicatoren.

6. Conclusies: koppeling theorie en empirie

6.1. De dominante beelden uit hoofdstuk 5 geaggregeerd

Wat is nu, gegeven de vele beelden in hoofdstuk 5, het geaggregeerde beeld?

De onderzochte regio's erkennen het belang van inzicht in prestaties, van de systeemverantwoordelijkheid van de minister en van verticale verantwoording. Gesprekspartners in de regio's erkennen ook het belang van prestatie-indicatoren, daar waar het gaat om de wettelijke vereisten.

Tegelijkertijd is er kritiek op, dan wel angst voor een mechanistisch gebruik van prestatie-indicatoren. Er is weinig steun voor het vervangen van huidige indicatoren door output- en outcome-indicatoren. Ten eerste omdat output- en outcome-prestaties in veiligheidsregio's zich niet laten vatten in een systeem van indicatoren. Ten tweede omdat door de lage frequentie en het unieke karakter van een ramp/crisis de zeggingskracht van indicatoren in de warme fase beperkt is. Ten derde omdat de relatie koud – warm in een systeem van prestatiemeting problematisch is. Ten vierde omdat er veel angst is dat indicatoren zullen worden gebruikt voor vormen van afrekenen – in de brede zin van het woord. Ranking wordt ook ervaren als een vorm van afrekenen. De angst voor afrekenen komt voort uit (1) het gegeven dat regio's zijn gerankt en (2) de onzekerheid over de rol die indicatoren uit de verticale verantwoordingsrelatie zullen spelen in horizontale verantwoordingsrelaties. Indicatoren hebben dus een gepercipieerde hoge impact.

Er worden dan ook diverse ongewenste 'perverse' effecten van prestatiemeting genoemd: afvinken en bureaucratie, de verhullende werking van prestatie-indicatoren – ze geven inzicht, maar laten ook aspecten buiten beschouwing – en een enkel voorbeeld van 'gaming the numbers'.

Veel beelden die ons zijn voorgehouden, hadden betrekking op de context van de veiligheidsregio's. De discussie over prestatie-indicatoren wordt in een context gevoerd die zich kenmerkt door:

- De verschillen die er zijn tussen veiligheidsregio's.
- De (gepercipieerde) spanning tussen verticale en horizontale verantwoording.
- De dynamiek van de regio's, die vaak nog eenheden-in-ontwikkeling zijn.

Voorts wijzen gesprekspartners in de regio's erop dat er reeds verschillende initiatieven bestaan om te leren, te evalueren en kwaliteit te borgen.

6.2. Waardering van de beelden

Hoe waarderen wij, als onderzoekers, de beelden uit de regio's? Dat is een lastige vraag. Op een denkbeeldig spectrum zijn er twee uitersten denkbaar. Het eerste: de regio's hebben helemaal gelijk. Het tweede: de regio's grossieren in drogredenen om hun verantwoordingsplicht te ontlopen.

De waardering van de beelden is ook lastig omdat er natuurlijk een functionele spanning is tussen de Inspectie VenJ en de veiligheidsregio's. Daar waar de regio's soms een negatief oordeel hebben (bijvoorbeeld over ranking), is vanuit het perspectief van de Inspectie een positief oordeel mogelijk – of andersom.

Om desondanks tot een waardering te komen, toetsen we hierna de beelden van de gesprekspartners uit de regio's aan de in hoofdstuk 4 gepresenteerde ontwerpeisen.

6.2.1. Meetbaarheid, toerekenbaarheid en relevantie

We laten alle overwegingen die we hierboven onder 'context' hebben benoemd (verschillen tussen regio's, verticale en horizontale verantwoording, dynamiek) eerst buiten beschouwing. We gaan ook nog niet in op de (gepercipieerde) impact van prestatiemeting, maar richten ons op de basale vraag: zijn de prestaties van een veiligheidsregio goed in indicatoren te vangen? Zijn ze meetbaar, toerekenbaar en relevant?

Om te beginnen gaan we in op de regelgeving uit de wet. Vervolgens behandelen we de mogelijkheid van output- en outcome-indicatoren.

Wettelijke vereisten

Het belang van het op orde hebben van de wettelijke vereisten uit de regelgeving is nauwelijks omstreden. De wettelijke vereisten zijn over het algemeen ook goed meetbaar en toe te rekenen aan de inspanningen van de veiligheidsregio's. Het is bijvoorbeeld relatief eenvoudig om vast te stellen of regio's over een beleidsplan, een crisisplan en een risicoprofiel beschikken. En regio's hebben zelf in de hand of ze deze zaken op orde hebben.

Edoch, voor elk van die eisen geldt wel dat het uiteindelijk gaat om de werkelijkheid achter die cijfers. We willen uiteindelijk weten wat er in werkelijkheid gebeurt, bijvoorbeeld met een crisisplan of een risicoprofiel. Hoe degelijk was het totstandkomingsproces? Hoe worden plannen en profielen levend gehouden, welke betekenis hebben die voor het denken en doen in de regio? Het enkele feit dat een indicator wordt gehaald of niet gehaald, zegt nog heel weinig. Zijn de opkomsttijden niet op orde? Dat kan desondanks betekenen dat een goede prestatie is geleverd, omdat de diensten eerst de tijd namen om informatie in te winnen over de situatie op de plek van de ramp. Zijn ze wel op orde? Dat kan betekenen dat er desondanks geen goede prestatie is geleverd, omdat de diensten niet de tijd hebben genomen om zich te informeren. Bovendien wordt in de huidige wijze van beoordelen de relatieve zwaarte van de overtreding niet meegewogen. Of een functionaris 5 minuten of 3 uur te laat verschijnt blijkt niet uit de indicator, maar maakt in de praktijk nogal een verschil. Daarmee worden de afrekenmechanismen te grofmazig.

Kortom: indicatoren die uitdrukking geven aan de wettelijke vereisten uit de wet zijn weliswaar goed meetbaar en toerekenbaar, maar ze worden pas relevant als er ruimte is voor kwalitatieve verdieping waarmee zicht wordt verkregen op de werkelijkheid achter de cijfers.

Output- en outcome-indicatoren

De vraag die in dit onderzoek voorligt, is of output- en outcome-indicatoren als passend alternatief kunnen dienen voor of als aanvulling kunnen dienen op het meten van de wettelijke vereisten. Het antwoord hierop is op het eerste gezicht positief: outcome-indicatoren – en in mindere mate ook output-indicatoren – zeggen meer over de prestatie van de veiligheidsregio dan het louter voldoen aan de wettelijke vereisten. Ze geven volgens de Inspectie, de NCTV en de veiligheidsregio's uitdrukking aan die zaken die er écht toe doen.

Wij constateren echter dat alle genoemde output- en outcome-indicatoren (zie tabel hoofdstuk empirie) aan één of beide van de volgende tekortkomingen lijden:

- Ze zijn niet goed eenduidig meetbaar. Zodra begrippen als ‘kwaliteit’ of ‘duidelijkheid’ vallen (die worden veel genoemd als onderdeel van output- en outcome-indicatoren), weten we een ding zeker: die bestaan uit veel, soms onderling conflicterende componenten, waarvan er op z’n minst een aantal niet meetbaar is. In een poging om prestaties toch meetbaar te maken, verliezen de indicatoren hun relevantie. We nemen de ‘kwaliteit van de crisiscommunicatie’ in de warme fase als voorbeeld. Een meetbare indicator zou kunnen zijn: het aantal optredens op televisie of in de geschreven pers. Echter, wat zegt deze indicator over de kwaliteit van de boodschap? En hoe normeer je het aantal optredens? Is veel beter? Er zijn ook situaties denkbaar dat weinig communiceren de voorkeur heeft. Natuurlijk is het mogelijk om een indicator toe te voegen om een beeld te krijgen van de kwaliteit. Een aanvullende meetbare indicator zou kunnen zijn: de tijdigheid van communicatie. Opnieuw is dan de vraag wat de indicator zegt over de kwaliteit. Is sneller altijd beter? Er zijn ook situaties denkbaar, waarin het verstandiger is om te wachten met de communicatie totdat er informatie van goede kwaliteit is.
- Ze zijn niet goed toerekenbaar. Voor veel outcome-indicatoren geldt dat ze niet eenduidig te relateren zijn aan het optreden van de regio. Prestaties zijn toerekenbaar als er een duidelijke causale keten is te construeren tussen een inspanning en een resultaat. Dat is bij veiligheidsregio’s bijzonder moeilijk. Van achter naar voren geredeneerd: er is een (goed of slecht) resultaat, dat kan eenduidig worden gerelateerd aan het optreden van de regio (is niet zo) en dat kan weer eenduidig worden gerelateerd aan de activiteiten in de koude fase (is ook niet zo). Voorbeeld: uiteraard is het aantal slachtoffers van een ramp relevant, maar dat cijfer op zich zegt nog niets over de prestaties van de regio. Er zijn immers allerlei interveniërende variabelen: je kunt weinig slachtoffers hebben, ondanks het slechte optreden van een regio. Het optreden van de regio in de warme fase, zegt nog niets over de relatie tussen dat optreden en de voorbereiding in de koude fase. Dus is altijd nadere analyse geboden, waarin kwalitatieve redeneringen ook een belangrijke rol spelen.

  Kortom: Output- en outcome-indicatoren zijn weliswaar relevant, maar zelden goed eenduidig meetbaar en toerekenbaar.

6.2.2. Context van veiligheidsregio’s

Nu gelden de bezwaren van het gebruik van outcome- en – in mindere mate – output-indicatoren voor veel (publieke) sectoren. In deze wereld van de veiligheidsregio’s zijn er echter enkele aanvullende bijzonderheden.

Ten eerste speelt in de veiligheidsregio’s de koud-warm dynamiek. Prestaties in de ‘warme’-fase (als zich daadwerkelijk een incident voordoet) hebben de meeste zeggingskracht – dat is wanneer het er écht toe doet – maar zijn uniek en komen weinig voor en zijn daardoor niet te vangen in indicatoren. Respondenten noemen als outcome-indicator ‘kwaliteit van nazorg’. De operationalisering daarvan hangt natuurlijk sterk af van de aard van de gebeurtenis. Prestaties in de ‘koude’-fase (als er niets aan de hand is en de nadruk ligt op zaken als opleiden en oefenen) zijn veelvoorkomend, maar zijn niet in output- en outcome-indicatoren te vangen.[27]

Ten tweede is sprake van een samenspel van mono en multi. Er zijn kolommen en er is de regio die toegevoegde waarde behoort te hebben ten opzichte van de kolommen. Veel van de activiteiten van de regio zijn relationeel van aard – overleg, samenwerking, kennisuitwisseling, afstemming, evalueren. Die kun je hooguit in procesindicatoren vangen, die ook nog bijzonder weinig zeggen.

---

[27] De realiteit kan in het beste geval worden benaderd door op een onverwacht moment een oefening te houden, zoals wordt beoogd met de systeemtest.

Kortom: De veiligheidsregio – als de context waarin prestatiemeting moet plaatsvinden – versterkt het problematische karakter van het gebruik van output- en outcome-indicatoren.

6.2.3. Hoge gepercipieerde impact

Uit de literatuur volgt dat als meetbaarheid en toerekenbaarheid problematisch zijn, zoals hier het geval is, dat de gepercipieerde impact van het prestatiemeetsysteem niet te hoog mag zijn. Perverse effecten zullen zich vooral manifesteren bij high impact systemen, die als oneerlijk of dom worden ervaren. Problematische toerekenbaarheid en meetbaarheid zullen het beeld van een oneerlijk en 'dom' systeem versterken.

Het gaat om een *gepercipieerd* high impact systeem. Die perceptie wordt gevoed door de onzekerheden die er zijn over het gebruik van het systeem. Hoe zal de minister de cijfers gebruiken? De Inspectie? Hoe zullen ze landen in de horizontale verantwoordingsrelatie? Welke rol zullen ze in de media spelen na een incident? Al die onzekerheden voeden de gedachte dat een systeem wel eens high impact zou kunnen zijn.

In de regio's zijn wij gestuit op veel weerstand tegen (extra) indicatoren. Hierbij zijn allerlei overwegingen genoemd: de variëteit tussen de regio's (die op gespannen voet staat met een eenduidig systeem), het relatief korte bestaan van de regio's (die dus nog in ontwikkeling zijn), de spanning tussen verticale en horizontale verantwoording (die ongelukkig interfereren). Ze leiden stuk voor stuk tot kritiek op het introduceren van nieuwe indicatoren.

Echter, ook als we de kritiek uit de regio buiten beschouwing laten, komen wij tot de conclusie dat de zoektocht naar output- en outcome-indicatoren niet heilzaam is. De belangrijkste overwegingen voor onze conclusies zijn immers dat de aard van de activiteiten van een veiligheidsregio zich slecht leent voor een systeem van output- en outcome-indicatoren (problematische meetbaarheid en toerekenbaarheid). En dat de gepercipieerde impact van het systeem hoog is. Dat is niet bevorderlijk voor het gezag van een systeem van prestatie-indicatoren.

De introductie van een systeem van output- en outcome-indicatoren zal tot een systeem leiden met een problematisch gezag. Als voor zo'n indicator geldt dat die of slecht is geoperationaliseerd (dus niet goed meetbaar is) of niet toerekenbaar is, dan zegt het wel of niet behalen van zo'n indicator weinig. Dan zijn er twee reacties mogelijk vanuit de regio.

- De constatering is dat een indicator wel of niet wordt gehaald, maar daar gaan de regio's schouderophalend aan voorbij – dat is wederom een schijnwerkelijkheid.
- De regio wil geen bestuurlijk gedoe en doet er alles aan om de indicator te halen – maar ook dat is uiteindelijk een schijnwerkelijkheid.

Kortom: De combinatie van problematische meetbaarheid en toerekenbaarheid en hoge gepercipieerde impact zorgt voor sterke prikkels voor pervertering en ritualisering. De sturende werking van een systeem bestaande uit output- en outcome-indicatoren zal gering zijn. Het zal – net als het huidige systeem – vrijwel zeker leiden tot 'afvinken' en een schijnwerkelijkheid creëren.

6.3. Conclusie

De conclusie is dat het ontwikkelen en implementeren van een stelsel van output- en outcome-indicatoren naar onze mening een weinig heilzame weg is. Er bestaat in de veiligheidsregio's weinig draagvlak voor (extra) indicatoren – en daar zijn argumenten voor. Wanneer wordt erkend dat het werk van de regio's zich niet in

een set indicatoren laat vangen, zou dit wel eens de ruimte kunnen bieden voor veel vruchtbaarder discussies over verantwoording en leren.

In de literatuur over prestatiesturing wordt vaak gesteld dat indicatoren (ook proces-indicatoren!) vooral een goed middel zijn om een gesprek op gang te krijgen over het presteren van, in dit geval, de regio's. Het wel of niet behalen van een indicator is niet het einde van een discussie (dat leidt inderdaad tot afvinken), maar het begin van een gefundeerde discussie.

7. Reflectie

7.1. Inleiding

Centraal in dit onderzoek staat de vraag in hoeverre en op welke wijze het mogelijk is om een stelsel van indicatoren te ontwikkelen waarmee een beeld kan worden verkregen van het presterend vermogen van veiligheidsregio's, evenals de vraag in hoeverre dit wenselijk is vanuit het perspectief van de betrokken actoren.

In het vorige hoofdstuk hebben we geconcludeerd dat het – bezien vanuit de theorie en de praktijk – niet mogelijk en wenselijk is om met één of enkele output- en/of outcome-indicatoren uitdrukking te geven aan het presterend vermogen van veiligheidsregio's.

Tegelijkertijd is de vraag van de Inspectie VenJ en de NCTV naar inzicht in het presterend vermogen van regio's volstrekt legitiem. Het antwoord is weliswaar niet gelegen in nieuwe output- of outcome-indicatoren, maar dat mag er niet toe leiden dat de vraag naar het presterend vermogen van regio's wordt gerelativeerd .

In dit laatste hoofdstuk reflecteren we daarom – op hoofdlijnen – op de vraag hoe binnen het bestaande stelsel meer rijke en gezaghebbende interactie met de veiligheidsregio's kan worden georganiseerd, zodat een beter inzicht ontstaat in het presterend vermogen van veiligheidsregio's. Hiermee pogen we een aanzet te geven voor een vruchtbare discussie over dit onderwerp.

We schetsen allereerst een aantal recente ontwikkelingen in het toezicht van de Inspectie VenJ. Vervolgens bieden we een aantal handreikingen voor het verder doorontwikkelen van het toezichtstelsel. Het doorontwikkelen van het toezichtstelsel vergt inspanningen van zowel het Rijk als van de regio's.

7.2. Ontwikkelingen in het toezicht op de veiligheidsregio's

Staat van de rampenbestrijding

De Inspectie VenJ rapporteert driejaarlijks[28] in de 'Staat van de rampenbestrijding' over de operationele prestaties op het terrein van de rampenbestrijding, bezien vanuit de vereisten uit de Wet en het Besluit veiligheidsregio's. Daarnaast wordt in de Staat op basis van dezelfde regelgeving ingegaan op planvorming en samenwerkingsafspraken. De laatste 'Staat van de rampenbestrijding' dateert uit 2013.[29] Hierbij is gebruik gemaakt van het 'Toetsingskader Staat van de rampenbestrijding'.[30]

Zelfevaluatie-instrument

De Inspectie VenJ wil de eigen verantwoordelijkheid van regio's stimuleren en de toezichtslast waar mogelijk beperken voor zowel veiligheidsregio's als de Inspectie zelf. Systeemoefeningen en grote incidenten (GRIP 3 en 4) moeten door de veiligheidsregio's zelf geëvalueerd worden. De Inspectie VenJ heeft hiervoor in 2012 een

---

[28] In 2010 verscheen de eerste Staat van de rampenbestrijding, in 2013 de tweede. De derde is voorzien in 2016.
[29] Staat van de Rampenbestrijding 2013, Onderzoek Rampenbestrijding op Orde periode maart 2010 - oktober 2012.
[30] Toetsingskader Staat van de Rampenbestrijding, Inspectie VenJ, 12 augustus 2013.

zelfevaluatie-instrument ontwikkeld.[31] Voor de oefeningen stelt de veiligheidsregio, op basis van het wettelijke kader, de verbeterpunten uit de 'Staat' en haar eigen ambities oefendoelen vast.

Protocol toezicht op de veiligheidsregio's

In februari 2015 is het 'Protocol toezicht op de veiligheidsregio's' geactualiseerd.[32] Het protocol is erop gericht om de taken en rollen van de Inspectie VenJ en de commissarissen van de Koning in het proces van toezicht op de taakuitvoering in de veiligheidsregio's goed op elkaar af te stemmen. Het protocol bevat afspraken voor de koude fase. In het protocol is onder meer een zogeheten toezichtsladder opgenomen, waarin sprake is van een oplopende mate van dwang:

- Wanneer de Inspectie VenJ tekortkomingen constateert in de wijze waarop een veiligheidsregio haar taken uitvoert, maakt de Inspectie (verbeter)afspraken met het bestuur van de veiligheidsregio en stelt de termijn vast waarbinnen deze moeten worden gerealiseerd.
- De Inspectie VenJ monitort de (verbeter)afspraken en informeert het bestuur van de veiligheidsregio als afspraken naar het oordeel van de Inspectie onvoldoende worden nagekomen.
- Als blijkt dat de uitvoering van de (verbeter)afspraken onvoldoende uitzicht biedt op verbetering, vindt er overleg plaats tussen de commissaris van de Koning en het hoofd van de Inspectie VenJ. De Inspectie VenJ ziet erop toe dat eventuele vervolgafspraken worden nagekomen door de veiligheidsregio.
- Wanneer de nadere afspraken niet of onvoldoende worden uitgevoerd, kan de commissaris van de Koning op grond van de Wvr en zijn ambtsinstructie het bestuur van de veiligheidsregio een aanwijzing geven. De minister dient in te stemmen met het geven van een aanwijzing. Het bestuur van de veiligheidsregio is verplicht een aanwijzing op te volgen.

7.3. Van 'afvinken' naar kritische interactie

7.3.1. Meer nadruk op het gesprek, minder op de indicator

Ondanks het gegeven dat de Inspectie VenJ het toezicht op de taakuitvoering in de veiligheidsregio's in de afgelopen jaren heeft doorontwikkeld (zie paragraaf 7.2), blijft in de regio's het beeld bestaan dat het toezicht te veel is gebaseerd op 'afvinken': het biedt volgens hen beperkt inzicht in het presterend vermogen van de veiligheidsregio's. De Inspectie VenJ en de NCTV erkennen de beperkingen van de huidige systematiek.

In de conclusies hebben we aangegeven dat het ontwikkelen en introduceren van output- en outcome-indicatoren zeer waarschijnlijk ook tot afvinken zal leiden. De aard van de prestaties van de regio leent zich niet voor indicatoren die je meet, maar meer voor een goede agenda waarover je spreekt. Als gebruik wordt gemaakt van indicatoren, zijn ze niet het einde van de discussie, maar het begin van de discussie. Nog weer anders geformuleerd: de vraag naar de output en outcome van een regio is volstrekt legitiem; je beantwoordt de vraag alleen niet door (nieuwe) prestatie-indicatoren te ontwikkelen.

Het gaat dus meer om het gesprek, minder om de indicator. Natuurlijk vinden er al gesprekken plaats tussen zowel de Inspectie en de NCTV als de regio's, maar deze leiden op dit moment nog niet tot het gewenste inzicht in het presterend vermogen van de regio's. Ook is er al in een aantal regio's geëxperimenteerd met visitaties. Deze visitatiesystematiek is nog in ontwikkeling.

[31] Zelfevaluatie-instrument, Handreiking evaluatie systeemoefeningen en GRIP 3/4-incidenten, Inspectie VenJ, augustus 2012.
[32] Protocol toezicht op de veiligheidsregio's, Afspraken tussen de commissarissen van de Koning en het hoofd van de Inspectie Veiligheid en Justitie, 26 februari 2015.

In deze reflectie doen we een aantal handreikingen voor het voeren van het goede gesprek. Allereerst generiek: wat zijn dimensies van zulke gesprekken die kunnen helpen om deze goed te voeren? Vervolgens doen we een aantal handreikingen voor kritische interactie en de vormgeving van zo'n gesprek.

De verantwoordelijkheid voor het vormgeven en op gang brengen van deze gesprekken ligt niet alleen bij de Inspectie en de NCTV. De vruchtbaarheid van de interactie is afhankelijk van de bereidheid van regio's om inzicht te geven in hun presterend vermogen – niet alleen in wat goed gaat, maar ook in wat beter kan, wat onzekerheden en zorgen zijn.

7.3.2. Dimensies van interactie

Hierna beschrijven we vijf dimensies waarop keuzes gemaakt kunnen worden, ten einde te komen tot kritische interactie tussen het ministerie van VenJ (Inspectie en NCTV) en regio, en beter inzicht te verkrijgen in het presterend vermogen van de veiligheidsregio's.

- *Verantwoorden of leren.* 'Slechts' leren of worden de resultaten van de gesprekken gebruikt voor verticale en horizontale verantwoording?
- *Retrospectief of prospectief.* Gaat het vooral om de past performance (die een indicatie is voor de kwaliteit van toekomstig handelen) of kijken we ook vooruit?
- *Individueel of vergelijkend.* Kijkt de Inspectie naar de individuele regio en maakt ze met de individuele regio afspraken of kan ze de output van gesprekken ook gebruiken voor vergelijkingen?
- *Besloten of openbaar?* Vindt het gesprek plaats tussen Inspectie en regio en zijn de uitkomsten niet voor derden beschikbaar? Of zijn ze dat wel?
- *Individueel oordeel of ranking?* Worden de oordelen van de Inspectie gebruikt voor rankings? Of velt ze alleen oordelen over individuele regio's?

7.3.3. Handreikingen voor kritische interactie

Mede op grond van de interviews die we hebben gevoerd, zou dit de essentie van de doorontwikkeling van het toezicht kunnen zijn:

De aandacht verschuift van een focus op nieuwe output- en outcome-indicatoren naar het 'gesprek' of de 'kritische interactie' tussen de ministerie van VenJ (Inspectie en NCTV) en regio en tussen de regio's onderling. Dat biedt de mogelijkheid om tot de essentie van waar het om gaat in de veiligheidsregio's door te dringen.

Combineren van kwalitatieve en kwantitatieve beelden

Een eerste aanname achter dit pleidooi voor een gesprek is dat het combineren van kwalitatieve en kwantitatieve beelden tot een beter inzicht leidt in de prestaties van een veiligheidsregio. Veel van wat de veiligheidsregio doet, is namelijk niet te vatten in alleen een indicator.

Voorbeeld: Crisisoverleg
Als bijvoorbeeld wordt gekeken naar een crisisoverleg, gaat het natuurlijk om de kwaliteit van dat overleg. Is de voorzitter competent? Weet deze het evenwicht te bewaren tussen goed luisteren en besluitvaardigheid? Als de voorzitter niet beschikbaar is, is er dan een competente plaatsvervangende voorzitter? Hoe is het team samengesteld en op basis van welke afwegingen is dat gedaan? Heeft het team een vaste samenstelling, kennen de spelers elkaar voldoende? Komen de relevante perspectieven aan de orde voordat er wordt besloten? Kunnen de voorzitter en de leden met de tijdsdruk omgaan?

Een aantal van de in het voorbeeld genoemde aspecten is niet te vangen in een indicator, bijvoorbeeld die van de competentie van een voorzitter. Dus is het zaak daarover het gesprek aan te gaan, bijvoorbeeld door de spelregel te hanteren dat de voorzitter feedback krijgt van deelnemers, en dat de Inspectie iets zegt over (a) de competentie van de voorzitter en (b) de kwaliteit van de feedback.

Theoretisch is een aantal aspecten wel te vangen in een indicator, bijvoorbeeld de vaste samenstelling van een team. Maar al snel komen daar toch weer lastige vragen op over de operationalisering daarvan. Als iemand zich laat vervangen door een plaatsvervanger, is dat dan een vaste samenstelling? En als in de vaste samenstelling toch twee zwakke spelers zitten, wat heb je dan aan een goede score op 'vaste samenstelling'?

Confronteren van perspectieven

Een tweede belangrijke aanname achter het pleidooi voor een gesprek is dat zowel de Inspectie als de regio's een legitiem perspectief kunnen hebben op wat een goede prestatie is, en dat die perspectieven wel eens kunnen schuren. Dan is het noodzakelijk om die perspectieven met elkaar te confronteren.

Ons zijn voorbeelden genoemd die – enigszins gestileerd – passen in kwadrant C van figuur 4. De regio vindt dat ze bij een incident goed heeft gehandeld – en weet zich gesteund door de cijfers. De Inspectie heeft kritiek op de afhandeling van het incident. Die twee beelden kunnen allebei juist zijn en dus gaat het erom dat je met elkaar het gesprek aangaat, om die beelden met elkaar te confronteren en daar lessen uit te trekken.

	Regio: goede prestatie	**Regio: slechte prestatie**
Inspectie: goede prestatie	A	B
Inspectie: slechte prestatie	C	D

Figuur 4. Verschillende perspectieven op prestaties.

> **Voorbeeld: Hulpverlening**
> Stel we nemen een grote uitslaande brand met een groot aantal gewonden. De regio oordeelt positief over haar eigen functioneren en baseert zich hierbij op een enquête waaruit blijkt dat de slachtoffers tevreden zijn over de hulpverlening door de hulpdiensten. De Inspectie is echter kritisch. Uit onderzoek blijkt namelijk dat het proces van triage (het classificeren van gewonden naar de ernst van hun verwondingen) niet goed is verlopen en dat in naburige ziekenhuizen meer capaciteit is vrijgemaakt voor het opvangen van gewonden dan nodig bleek te zijn.

Ons zijn ook voorbeelden genoemd waarbij de uitkomst van een indicator niet in lijn is met het oordeel van de betrokken professional(s): kwadrant B of C in figuur 5.

	Prestatiemeting: goede prestatie	**Prestatiemeting: slechte prestatie**
Professionele oordeel: goede prestatie	A	B
Professionele oordeel: slechte prestatie	C	D

Figuur 5. Uitkomst van prestatiemeting versus het professionele oordeel.

Voorbeeld: Aanrijtijden

Denk hier aan het voorbeeld van de aanrijtijden. In kwadrant A en D leiden professionele oordelen en het prestatiecijfer tot eenzelfde conclusie. In de andere kwadranten is dat niet zo. In kwadrant B kan de situatie zijn dat de aanrijdtijd niet is gehaald, maar dat het professionele oordeel over de prestatie toch positief is, omdat er terecht eerst wat tijd is genomen om informatie in te winnen over de toestand te plekke. In kwadrant C doet zich het omgekeerde voor: de aanrijtijd is binnen de norm, maar professioneel bezien is het geen goede prestatie – had er tijd genomen moeten worden om een beeld te krijgen van de situatie ter plekke. Voor kwadrant B en C geldt: alleen een goed gesprek kan tot zinvolle beelden leiden.

Door in het gesprek tussen Inspectie en regio indicatoren te combineren met het oordeel van betrokken professionals ontstaat een rijker beeld van de prestaties van een veiligheidsregio.

Dat gesprek is een intelligent vormgegeven gesprek en niet vrijblijvend

'Gesprek' of 'kritische interactie' klinkt uiteraard erg vrijblijvend. Dat mag het niet zijn. Het moet juist een gesprek zijn dat niet vrijblijvend is.

- Er is een duidelijk inhoudelijk format voor de inhoud van dat gesprek. Dat format geeft aan welke aspecten aan de orde moeten komen en geeft per aspect aan wat de intelligente vragen zijn.
- Er is een duidelijke procesgang, gebaseerd op vooraf afgesproken spelregels: het gesprek wordt met een zekere regelmaat gevoerd, er zijn bepaalde spelregels over wie aan tafel zit, wat de agenda is en hoe het gesprek wordt vormgegeven.
- De spanning tussen leren en verantwoorden wordt geadresseerd. Er worden spelregels gemaakt over de manier waarop met deze spanning wordt omgegaan.

Over het inhoudelijke format

Als het gaat om het inhoudelijke format voor dat gesprek, kan bijvoorbeeld worden gedacht aan elementen van het huidige toetsingskader van de Inspectie. Elk van de elementen van het toetsingskader leent zich voor een aantal parameters of slimme vragen of indicatoren (die dan niet altijd kwantificeerbaar zijn), die kunnen bijdragen tot een rijker gesprek.

Voorbeeld: Risicoprofiel

Neem bijvoorbeeld het vereiste dat er een risicoprofiel moet zijn. Hierna volgen enkele voorbeelden van normen uit het huidige toetsingskader van de Inspectie VenJ ten aanzien van deze vereiste:[33]

- Is in het risicoprofiel een overzicht opgenomen van risicovolle situaties binnen de eigen veiligheidsregio die tot een brand, ramp of crisis kunnen leiden? Ja / nee

---

[33] Andere normen uit het toetsingskader met betrekking tot het risicoprofiel zijn:

- Is het beleidsplan van de veiligheidsregio mede gebaseerd op het vastgestelde risicoprofiel? Ja / nee
- Is in het risicoprofiel een overzicht opgenomen van de soorten branden, rampen of crises die zich in bovengenoemde regio kunnen voordoen? Ja / nee
- Zijn de gemeenteraden van de deelnemende gemeenten gevraagd om hun wensen kenbaar te maken? Ja / nee
- Zijn het regionale college, de besturen van de betrokken waterschappen en de door Onze andere Ministers daartoe aanwezen functionarissen uitgenodigd hun zienswijze ter zaken kenbaar te maken? Ja / nee
- Is er minimaal eenmaal per jaar gezamenlijk overleg tussen de veiligheidsregio en de bij mogelijke rampen en crises in de regio betrokken partijen? Ja / nee
- Is in het risicoprofiel een overzicht opgenomen van risicovolle landsgrensoverschrijdende situaties? Ja / nee
  - Heeft de regio risicovolle landsgrensoverschrijdende situaties gesignaleerd? Ja / nee
  - Heeft de regio risicovolle landsgrensoverschrijdende situaties uitgewerkt? Ja / nee
- Is in het risicoprofiel een informatieoverzicht opgenomen van bovenregionale risico's?
  - M.b.t. pandemieën? Ja / nee
  - M.b.t. overstromingen? Ja / nee
  - M.b.t. vitale infrastructuur? Ja / nee

- Is in het risicoprofiel een analyse opgenomen met betrekking tot de weging en inschatting van de gevolgen van branden, rampen en crises? Ja / nee
- Heeft het bestuur van de veiligheidsregio het risicoprofiel vastgesteld na overleg met de raden van de deelnemende gemeenten? Ja / nee
- Worden de multidisciplinaire oefeningen aangepast op het risicoprofiel van de regio? Ja / nee

Op zichzelf wezenlijk belangrijke elementen, maar zonder kwalitatieve verdieping weinigzeggend. Immers, dat het risicoprofiel een overzicht van risicovolle situaties bevat zegt niets over de accuraatheid van die lijst. En dat gemeenteraden zijn geconsulteerd zegt niets over de diepgang waarmee dit is gebeurd.

In een gesprek kan deze verdieping worden aangebracht. Hierna volgen voorbeelden van intelligente vragen over zo'n risicoprofiel, die in een gespreksformat opgenomen zouden kunnen worden:

- Welke data zijn gebruikt? Zijn er concurrerende data die gebruikt hadden kunnen worden? Hoe weet u zeker dat de gebruikte data tot het goede profiel leiden?
- Zijn er nieuwe ontwikkelingen, die van invloed zijn op het profiel? Hoe draagt u er zorg voor dat u die ontwikkelingen volgt en indien nodig, vertaalt in het risicoprofiel?
- Hoe heeft u de verschillende componenten van een risicoprofiel gewogen? Was een andere weging mogelijk geweest, en had die tot een ander risicoprofiel kunnen leiden?

Voor deze en dergelijke vragen geldt dat ze relevant lijken om eens goed door te vragen op de robuustheid van een risicoprofiel. Misschien vinden de regio's dit niet de goede vragen, maar dan moeten ze op z'n minst kunnen aangeven waarom dat zo is. Het zijn ook vragen die je in een gesprek aan de orde moet stellen, zodat een levendig vraag- en antwoordspel kan ontstaan. Worden deze vragen op papier gesteld, dan ontaarden ze al snel weer in een bureaucratische exercitie.

Voorbeeld: Convenanten met vitale partners
Een ander bijvoorbeeld betreft de vereiste dat de veiligheidsregio afspraken heeft gemaakt met vitale partners. Hierna volgen enkele voorbeelden van normen uit het huidige toetsingskader van de Inspectie VenJ ten aanzien van de samenwerking met vitale partners:

- Heeft de veiligheidsregio convenanten afgesloten met betrekking tot de volgende onderwerpen: Energie / gas? Drinkwater? Telecommunicatie? Spoor? Rijkswaterstaat / waterschappen? Ja / nee
- Bestaat het Commando Plaats Incident uit alle functionarissen zoals omschreven in artikel 2.1.2 lid 1? Heeft aan het CoPI een liaison Vitaal – CoPI deelgenomen? Ja / nee

Wederom belangrijke elementen, maar zonder kwalitatieve verdieping weinigzeggend. Immers, wat zegt het hebben van een convenant over de bekendheid van partners met elkaars processen en procedures? En dat er een liaison aanwezig was bij het incident zegt nog niets over de kwaliteit van het optreden van deze persoon.

In een gesprek kan diepgang worden aangebracht door – bijvoorbeeld – de volgende vragen te stellen:

- Wanneer zijn de convenanten voor het laatst geactualiseerd? Hoe wordt gewaarborgd dat het convenant 'leeft' bij zowel de regio als de partner?
- Wat zijn voor u als bestuurder de belangrijkste issues in die convenanten – waar zitten de belangrijkste zorgen? Hoe zorgt u er voor dat u die convenanten levendig houdt?
- Wanneer zijn de afspraken in het convenant voor het laatst aan bod gekomen in een oefening? Hoe verliep de samenwerking in die oefening? Wat is hiervan geleerd?

- Beschikt de veiligheidsregio over vaste contactpersonen bij de vitale partners? Hoe vertrouwd zijn de contactpersonen van de regio en de partners met elkaar?
- Hebben er recent incidenten plaatsgevonden waarbij vitale partners betrokken waren? Hoe verliep de samenwerking, zowel bestuurlijk als operationeel? Hoe vervulde de liaison Vitaal – CoPI zijn rol?

Over de spelregels

Interactie vergt ook een set spelregels, die het gesprek voorspelbaar maken. Wij kunnen uiteraard geen set spelregels aanbieden. Dat valt buiten de scope van dit onderzoek en, belangrijker, zulke spelregels vergen eigenaarschap van de betrokken spelers. Spelregels moeten dus door deze spelers worden gemaakt. We kunnen wel een aantal onderwerpen aanreiken die spelregels vergen.

- Er moet een inhoudelijke agenda zijn: welke onderwerpen komen aan bod in het gesprek tussen Rijk en regio?
- Er moet duidelijk zijn wie de gesprekken voert en er moet ook een zekere symmetrie zijn. Bestuurlijke gesprekspartner uit de regio vergen wellicht een andere vertegenwoordiging vanuit het Rijk dan operationele gesprekspartners.
- Met welke intensiteit worden gesprekken gevoerd tussen Rijk en regio?
- Hoe wordt gerapporteerd over de uitkomsten van gesprekken? Wat komt wel/niet in het verslag? Aan wie worden verslagen van gesprekken beschikbaar gesteld? Zie hierover ook de volgende paragraaf.
- Wanneer, hoe en op welke wijze wordt de minister geïnformeerd over de uitkomsten? Als uit het gesprek blijkt dat prestaties onder de maat zijn, wanneer wordt de minister daar dan van op de hoogte gesteld? Zie hierover ook de volgende paragraaf.
- Hoe worden de gesprekken ingepast in de bestaande gesprekscyclus van de NCTV en de Inspectie VenJ met de veiligheidsregio's?
- Hoe verhoudt de rapportage de 'Staat van de rampenbestrijding' zich tot het gesprek tussen Rijk en regio. Vinden gesprekken plaats naar aanleiding van de rapportage? Of gaat het gesprek vooraf aan de rapportage? Voor beide opties valt iets te zeggen.
- Hoe kan het gesprek tussen Rijk en regio zo goed mogelijk aansluiten bij bestaande horizontale verantwoordingsmechanismen (tussen regio en lokaal bestuur)?
- Hoe verhouden andere initiatieven als Aristoteles en Cicero zich tot het gesprek tussen Rijk en regio?
- Hoe kan het gesprek tussen Rijk en regio zo goed mogelijk aansluiten bij de visitatiesystematiek die in ontwikkeling is? Het meest voor de hand ligt om de visitatie te gebruiken voor horizontale verantwoording tussen regio's. Visitatierapporten kunnen vervolgens uiteraard een rol spelen in het gesprek van de NCTV en de Inspectie met de regio.

Nogmaals, de gesprekspartners zullen moeten afspreken hoe ze deze spelregels concreet maken.

Over de interventieladder, leren en verantwoorden

Leren vergt een zekere veiligheid in een besloten omgeving – het vertrouwen tussen Inspectie en regio's dat vrijuit kan worden gesproken, dat onzekerheden en dilemma's kunnen worden gedeeld. Maar wanneer het accent heel sterk op leren ligt, is er het risico dat het gesprek te vrijblijvend wordt – je wilt als regio leren, je koopt tijd voor verbetering, na een jaar is iets nog niet op orde, je beroept je op nieuwe omstandigheden, je hebt weer tijd nodig om te leren, et cetera.

Wie zich verantwoordt, verlaat de veiligheid van de besloten omgeving en komt in het politieke debat terecht. Verantwoorden vergt duidelijkheid: je geeft aan wat wel en niet op orde is, neemt daar publiek verantwoordelijkheid voor, inclusief voor de te realiseren verbeteringen. Wanneer het gesprek over verantwoording gaat, is het minder veilig en bestaat de kans dat er minder op tafel komt.

Het periodieke gesprek is op leren gericht, maar er komt natuurlijk ook informatie op tafel die door de Inspectie in de verantwoording kan of moet worden gebruikt. Hoe met deze spanning om te gaan?

We hebben daarop een eenvoudig antwoord: ook daarover moeten Inspectie en regio afspraken of spelregels maken. Op hoofdlijnen kunnen die er als volgt uitzien:

- Over de wettelijke vereisten wordt verantwoording afgelegd;
- voor het overige is het gesprek gericht op leren;
- de verantwoording over de wettelijke vereisten wordt publiek gemaakt;
- de belangrijkste afspraken over leren worden tussen Inspectie en regio vastgelegd;
- in vervolggesprekken wordt op die afspraken teruggekomen;
- als de Inspectie van mening is dat de regio die afspraken in onvoldoende mate nakomt, geeft ze aan dat die onderdeel moeten worden van de verantwoordingsrelatie – dus dat die bij onvoldoende voortgang publiek worden gemaakt;
- persisteert de regio in niet-nakoming, dan wordt dat gegeven onderdeel van de verantwoording;
- het omgekeerde geldt ook: daar waar de regio de afspraken goed nakomt en goed presteert, kan dat onderdeel worden van de verantwoordingsrelatie – en kan die informatie dus publiek worden gemaakt.

De systematiek hierachter is gebaseerd op het bekende idee van de interventieladder (Ayres en Braithwate 1992). Toezicht is gericht op leren en verbeteren, maar daar waar leren en verbeteren vervolgens geen gestalte krijgt, worden meer dwingende middelen ingezet.

Regio's zullen uiteraard moeten accepteren dat dit soort gesprekken niet afdoen aan de bevoegdheden en verantwoordelijkheid van de Inspectie. Die kan, gegeven die bevoegdheden en verantwoordelijkheid, altijd interveniëren als de wet dat, bijvoorbeeld, vereist. Dat is nu eenmaal de kernwaarde van een inspectie – en die kan nooit opzij worden gezet.

7.4. Tot slot

Van indicatoren naar een gesprek: het lijkt de overgang van iets dat hard en uitdagend is, naar iets dat zacht en (te) gemakkelijk is. Het tegenovergestelde is waar. Indicatoren zijn bij nadere beschouwing vaak boterzacht. En een gesprek kan een harde confrontatie zijn. Bovendien, een gesprek vergt ook wat van de Inspectie en de regio. Indicatoren scheppen afstand: je meet of een indicator is behaald en komt tot je conclusie. Daarvoor is, bij wijze van spreken, geen enkel onderling contact nodig. Als de aandacht verschuift naar het gesprek, is interactie geboden. De regio moet zich intelligenter verantwoorden. De Inspectie moet intelligenter kijken. De aandacht moet minder gericht zijn op indicatoren en meer op een gesprek. Dat gesprek mag niet vrijblijvend zijn: het doet niet af aan de verplichting te voldoen aan de wettelijke eisen. Het is gericht op leren en verbeteren, maar ook op verantwoorden. Voor de spanning die er nu eenmaal is tussen leren en verantwoorden, zijn spelregels nodig. Informatie uit de gesprekken kan dan worden gebruikt voor leren en verbeteren waar het kan en verantwoorden waar het moet.

Bijlagen

Literatuur

- Algemene Rekenkamer (2014) *Zicht overheden op beschermen burgers en bedrijven*. Den Haag: Algemene Rekenkamer.
- Atkinson, R. (1999) Project management: cost, time and quality, two best guesses and a phenomenon, it's time to accept other success criteria. *International Journal of Project Management* 17 (6): 337-342.
- Ayres, I. en J. Braithwate (1992), *Responsive Regulation*, Oxford: Oxford University Press
- Behn, R.D. (2003) Why measure performance? Different purposes require different measures. *Public Administration Review* 63 (5): 586-606.
- Berenschot (2009) *Eindrapportage Aristoteles. Prestatiemeting en –verantwoording in de veiligheidsregio's.* Utrecht: Berenschot.
- Bouckaert, G. (2006) Prestaties en prestatiemanagement in de publieke sector. *Tijdschrift voor Economie en Management* (3): 237-265.
- Bouckaert, G. en Halligan, H. (2007) Comparing Performance across Public Sectors*. Paper for presentation to Study Group on Performance in the Public Sector, Conference of European Group of Public Administration, Madrid, 19-22 September 2007.*
- Brink, G. van den, Jansen T. en Pessers, D. (2005) *Beroepszeer. Waarom Nederland niet goed werkt.* Amsterdam: Boom.
- Bruijn, H. de (2004) Prestatiemeting in de publieke sector. In: Knaap, P. van der, en Schilder, A. (red.) *Resultaatgericht sturen en evalueren. Nieuwe perspectieven op beleidsinstrumenten en beleidsevaluatie*. Den Haag: Sdu Uitgevers, pp. 195-203.
- Bruijn, H. de (2007) *Managing Performance in the Public Sector*. London: Routledge.
- Bruijn, H. de (2008) *Over management als probleem en als oplossing.* Den Haag: Academic Service.
- Dijkzeul, A., Boer, N.-I. en Mil, B. van (2008) Politiekorpsen: Aantoonbaar bijdragen aan een veiliger Nederland. In: Mil, B. van, Boer, N.-I., Velzen, G. van, Bruijn, H. de, Heuvelhof, E. ten, en Dijkzeul, A. (red.) *Prestaties afspreken doe je zo! Lessen over outputsturing uit acht sectoren.* Assen: Van Gorcum, pp. 21-32.
- Duiverman, O. (2014) *Benchmarking Performance of Offshore Construction Projects. A case study for creating a suitable benchmark for AJS offshore construction projects on the ONEgas gas assets in the Southern North Sea.* Master Thesis Delft: TU Delft.
- Eshuis, R.J.J. (2008) *Het recht in betere tijden. Over de werking van interventies ter versnelling van civiele procedures.* Dissertatie EUR.

- Externe Onderzoekscommissie (2006) *Een tekortschietend zorgproces*. Zwolle: Inspectie voor de Gezondheidszorg.
- Goddard, R.D., Wayne, K. en Woolfolk Hoy, A. (2000) Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal* 37 (2): 479-507.
- Hasaart, F. (2011) *Incentives in the Diagnosis Treatment Combination payment system for specialist medical care. A study about behavioural responses of medical specialists and hospitals in the Netherlands.* Maastricht: Datawyse Universitaire Pers Maastricht.
- Inspectie Openbare Orde en Veiligheid (2009) *Prestatie-indicatoren brandweer.* Den Haag: Inspectie Openbare Orde en Veiligheid.
- Inspectie Openbare Orde en Veiligheid (2011) *Brand Chemie-Pack Moerdijk: Een onderzoek naar de bestrijding van (de effecten van) het grootschalig incident.* Den Haag: Inspectie Openbare Orde en Veiligheid.
- Inspectie Veiligheid en Justitie (2013) *Staat van de Rampenbestrijding 2013: Onderzoek Rampenbestrijding op Orde periode maart 2010 - oktober 2012.* Den Haag: Inspectie Veiligheid en Justitie.
- Inspectie Veiligheid en Justitie (2013) *Toetsingskader staat van de rampenbestrijding*. Den Haag: Inspectie Veiligheid en Justitie.
- Inspectie Veiligheid en Justitie (2014) *Evaluatie GRIP-3 incident Mosterdgas Ede: Evaluatierapport.* Den Haag: Inspectie Veiligheid en Justitie.
- Inspectie Veiligheid en Justitie (2014) *Rapport Brand aan de Kelders te Leeuwarden 19 oktober 2013.* Den Haag: Inspectie Veiligheid en Justitie.
- Jochoms, M.P.C.M., Laan, F. van der, Landman, W., Nijmeijer, P.S. en Sey, A. (2006) *Op prestaties gericht. Over de gevolgen van prestatiesturing en prestatieconvenanten voor sturing en uitvoering van het politiewerk*. Apeldoorn: Politie & Wetenschap, Politieacademie.
- Kanis, A. (2014) *Kwaliteitzorg in de veiligheidsregio, systeem of succes?* Afstudeeronderzoek Thorbecke Academie in samenwerking met ministerie VenJ.
- Kerpershoek, E. (2015) *Medical professionals' responses to a DRG performance management system for hospital care in the Netherlands. Reinterpreting 'perverse effects', such as upcoding and patient selection, using arguments of professionalism and managerialism.* 's Hertogenbosch: Uitgeverij BOXPress.
- Korsten, A.F.A. (2007) *Prestatiemanagement: doet management ertoe? Resultaten van organisaties of eenheden daarbinnen.*
- Kreis, R. (2012) De verborgen kosten van ons zorgsysteem. Overheid weigert bureaucratie aan te pakken. *Medisch Contact* 67 (33-34): 1890-1892.
- Maarse, H. en Velden, M. van (2004) Nutteloze Informatie. Het ene beste ziekenhuis is het andere niet. *Medisch Contact* 59 (44): 1722-1724.

- Mil, B. (2008) *Prestaties afspreken doe je zo! Lessen over outputsturing uit acht sectoren*. Assen: Van Gorcum.
- Ministerie van Volksgezondheid, Welzijn en Sport, Plexus en BKB (2010) *Werken aan de Zorg. Kosten en kwaliteit*. Den Haag: Ministerie van Volksgezondheid, Welzijn en Sport.
- NVBR (2008) *CICERO. Voorbij de normen van nu. Een meerjarenvisie op kwaliteit en innovatie voor de brandweer*. Arnhem: Nederlandse Vereniging voor Brandweerzorg & Rampenbestrijding.
- NVBR (2010) *Ontwikkelen en (kwaliteit)sturing bij de Brandweer. Functioneel ontwerp ontwikkelstelsel Brandweer. Resultaat actualisering en voorstel verdere ontwikkeling Cicero.* Arnhem: Nederlandse Vereniging voor Brandweerzorg & Rampenbestrijding.
- Osborne, D. en Gaebler, T. (1992) *Reinventing Government*. Reading: Addison-Wesley.
- Pen, M. (2009) *Prestatiemeting van wetenschappelijk onderzoek. Een empirische studie naar de doorwerking van universitaire prestatiemeetsystemen.* Den Haag: Boom Uitgevers.
- Pen, M. (2011) De paradox van de rigide prestatiemeting. Over de doorwerking van prestatiemeting in de omgeving van de academische professional. M&O (1): 5-17.
- Pollitt, C. (2013) The logics of performance management. *Evaluation* 19 (4): 346-363.
- Propper, C. en Wilson, D. (2003) The Use and Usefulness of Performance Measures in the Public Sector. *Oxford Review of Economic Policy* 19 (2): 250-267.
- RemBrand (projectgroep). (2015) Brandveiligheid is coproductie. Brandweer Nederland.
- Ridderstrale, J. en Nordstrom, K. (2004) *Karaoke Capitalism: Managing for Mankind.* Londen: Financial Times/Prentice Hall.
- Ruijs, G. (2015) Kwaliteit en vergelijkbaarheid. Magazine nationale veiligheid en crisisbeheersing (1): 13.
- Schoenmaker, R. (2011) *De ingeslagen weg: Een empirisch onderzoek naar de dynamiek van de uitbesteding van onderhoud in de civiele infrastructuur.* NGInfra PhD Thesis Series on Infrastructures. Delft: Next Generation Infrastructures Foundation.
- Schoenmaker, R. (2012) *De ingeslagen weg: Een empirisch onderzoek naar de dynamiek van de uitbesteding van onderhoud in fysieke infrastructuur.* Dissertatie TU Delft.
- Sylvester, D.C., Rani, N.S.A. en Shaikh, J.M. (2012) Comparison between oil and gas companies and contractors against cost, time, quality and scope for project success in Miri, Sarawak, Malaysia. *African Journal of Business Management* 5 (11): 4337-4354.
- Thiel, S. van, Leeuw, F.L. (2002) The performance paradox in the public sector. *Public Performance and Management Review* 25 (3):267-281.
- TU Delft, NICIS Institute en KWINK Groep (2010) *Handboek prestatiemeting. Kwaliteit in kaart*. Delft: TU Delft.

- Veiligheidsregio IJsselland (2014) *Veiligheidsregio IJsselland slaat twee vliegen in één klap: Cicero en de A3 methodiek als basis voor verandering!* Zwolle: Veiligheidsregio IJsselland.

- Veldhuisen, A. van, Hagelstein, R., Voskamp, I. en Genderen, I. van (2013) *Evaluatie Wet veiligheidsregio's. GV188-3-271.* Utrecht: Andersson Elffers Felix.

- VIJFadvies (2011) *Rapportage doorlichting veiligheidsregio Limburg-Noord*. Amsterdam: VIJFadvies.

- Wilson, D., Croxson, B. en Atkinson, A. (2006) What Gets Measured Gets Done. *Policy Studies* 27 (2): 153-171.

- Zouridis, S. (2011) *Politieprestaties in kaart. Prestaties en het presterend vermogen van de Politie.* Tilburg: Tilburgse School voor Politiek en Bestuur, Universiteit van Tilburg.

- Zouridis, S., Bouckaert, G., Roy, P., Stroobants, J. van, Crompvoets, V., Janssen, L. en Peeters, R. (2014) *Politieproductiviteit. Triangulatie voor valide en betrouwbare productiviteitsmeting bij de politie.* Tilburg: Tilburg University – KU Leuven.

Gesprekspartners

Gesprekspartners verkennende gesprekken

Gesprekspartners	Functie
Paul Gelton	directeur Weerbaarheidsverhoging NCTV, ministerie van VenJ
Gertjan Bos	hoofd Inspectie VenJ
Antoin Scholten	dagelijks bestuur Veiligheidsberaad, burgemeester Venlo, voorzitter Veiligheidsregio Limburg-Noord
Nico van Mourik	voorzitter Raad van directeuren veiligheidsregio, directeur veiligheidsregio Midden en West Brabant
Elie van Strien	commandant brandweer Amsterdam-Amstelland
Gerrie Ruijs	directeur veiligheidsregio Zeeland, commandant brandweer Zeeland
Peter Bos	directeur veiligheidsregio Utrecht, directeur publieke gezondheid Utrecht
Egbert Jongsma	projectleider Algemene Rekenkamer

Gesprekspartners in drie veiligheidsregio

Veiligheidsregio Amsterdam-Amstelland		
Niveau	**Gesprekspartners**	**Functie**
Gesprek op bestuurlijk niveau	Mirjam van 't Veld	Burgemeester gemeente Amstelveen
	Bauke van Bourgondiën	Senior (beleids)adviseur openbare orde en veiligheid, RIEC Amsterdam-Amstelland
	Bastiaan van 't Hoog	Concerncontroller Brandweer
Gesprek op directie niveau	Elie van Strien	Commandant Brandweer
	Jan Woldman	Operationeel Directeur Publieke Gezondheid, GHOR-bureau
Gesprek op ambtelijk niveau	Monique Groenhuijzen	Bestuursadviseur gemeente Amsterdam/Veiligheidsregio
	Ria Logtenberg	Algemeen beleidsmedewerker GHOR-bureau
	Jan van Asperen	Beleidsmedewerker GHOR-bureau
	Nicole Voltman	Manager Operationele Informatie Brandweer

Veiligheidsregio IJsselland		
Niveau	**Gesprekspartners**	**Functie**
Gesprek op bestuurlijk niveau	Henk Jan Meijer	Burgemeester gemeente Zwolle; voorzitter Veiligheidsregio
	Henk Schreuders	Bestuurssecretaris Veiligheidsregio
Gesprek op directie niveau	Lieke Sievers	Regionaal commandant Brandweer; voorzitter Veiligheidsdirectie
	Henk Schreuders	Bestuurssecretaris Veiligheidsregio
Gesprek op ambtelijk niveau	Arjen Schepers	Procesmanager Repressie; Commandant Cluster Midden
	Richard Kamphuis	Hoofd Veiligheidsbureau & CIO
	Annelies Grooten	Beleidsmedewerker Veiligheidsregio
	Henk Schreuders	Bestuurssecretaris Veiligheidsregio

Veiligheidsregio Gelderland-Zuid		
Niveau	**Gesprekspartners**	**Functie**
Gesprek op bestuurlijk niveau	Hans Verheijen	Burgemeester gemeente Wijchen; Lid algemeen bestuur Veiligheidsregio
Gesprek op directie niveau	Marijke van Veen	Directeur Veiligheidsregio
Gesprek op ambtelijk niveau	Marcel Meeuse	Hoofd Veiligheidsbureau
	Roland Pricken	Clustercommandant Brandweer
	Fonx Merkx	Beleidsadviseur Kwaliteit
	Frans Lischer	Hoofd GHOR

Gesprekspartners toetsingsbijeenkomst en toetsende gesprekken

Gesprekspartners	Functie
Paul Gelton	Directeur Weerbaarheidsverhoging NCTV, ministerie van VenJ
Gertjan Bos	Hoofd Inspectie VenJ

Toetsingsbijeenkomst overig regio's		
Gesprekspartner	Organisatie	Functie
Marieke Henneman	GHOR Haaglanden	Senior beleidsmedewerker/ plv. hoofd
Robert Jan Schouwerwou	GHOR Utrecht	Plv. directeur
Karin Dieleman	Veiligheidsregio Zeeland	Concerncontroller, lid Netwerk Finance & Control
Loek Smolenaars	Veiligheidsregio Limburg-Noord	Medewerker Kwaliteitszorg/ Kennisregisseur
Ernst-Jan Villerius	Veiligheidsregio Rotterdam Rijnmond	Senior controller/coördinator
Henk Meijer	Veiligheidsregio Hollands Midden	Directeur
Robert Zweegman	Brandweer Hollands Midden	Medewerker Integraal Beleid
Jacqueline Barnasconi	Veiligheids- en Gezondheidsregio Gelderland-Midden	Senior adviseur kwaliteitszorg
Ineke Vroling	Veiligheidsregio Drenthe	Adviseur kwaliteitszorg
Sietske Houkes	Veiligheidsregio Noord-Holland Noord	Procesexpert kwaliteit en procesmanagement
Maarten Dewachter	Veiligheidsregio Midden- en West-Brabant	Adjunct coördinerend gemeentesecretaris
Erik Kamphuis	Veiligheidsregio Noord-Holland Noord	Interimmanager / adviseur